



WICC²⁰₂₂
M E N D O Z A



LIBRO DE ACTAS

XXIV Workshop de Investigadores en Ciencias de la Computación
Mendoza, Abril de 2022

Universidad Champagnat - Red de Universidades de Carreras de Informática RedUNCI

Universidad Champagnat
Libro de Actas WICC 2022. -
La edición - Godoy Cruz : FUSMA Ediciones, 2022.
Libro digital, PDF
Archivo Digital: descarga

ISBN 978-987-48222-3-9

1. Computación. I. Título
CDD 005

Comité Académico

Universidad de Buenos Aires (Ciencias Exactas)
Ceria, Santiago

Universidad de Buenos Aires (Ingeniería)
Echeverría, Adriana

Universidad Nacional de La Plata
Pesado, Patricia

Universidad Nacional del Sur
Rueda, Sonia

Universidad Nacional de San Luis
Printista, Marcela

Universidad Nacional del Centro de la Provincia de Buenos Aires
Aciti, Claudio

Universidad Nacional del Comahue
Grosso, Guillermo

Universidad Nacional de La Matanza
Eterovic, Jorge

Universidad Nacional de La Pampa
Alfonso, Hugo

Universidad Nacional de Tierra del Fuego
Korembli, Gabriel

Universidad Nacional de Salta
Gil, Gustavo

Universidad Nacional de la Patagonia Austral
Lasso, Marta

Universidad Nacional de San Juan
Rodríguez, Nelson

Universidad Nacional de Rosario
Casali, Ana

Universidad Nacional de Misiones
Caballero, Sergio

Universidad Nacional del Noroeste de la Provincia de Buenos Aires
Russo, Claudia

Universidad Nacional de Chilecito
Carmona, Fernanda

Universidad Nacional de Lanús
Azcurra, Diego

Universidad Nacional de Santiago del Estero
Figuerola, Liliana

Escuela Superior del Ejército
Arroyo Arzubi, Alejandro

Universidad Nacional del Litoral
Loyarte, Horacio

Universidad Nacional de Río IV
Arroyo, Marcelo

Universidad Nacional de Córdoba
Fridlender, Daniel

Universidad Nacional de Jujuy
Herrera Cогnetta, Analía

Universidad Nacional de Río Negro
Vivas, Luis

Universidad Nacional de Villa María
Prato, Laura

Universidad Nacional de Entre Ríos

Mengarelli, José Luis

Universidad Nacional de la Patagonia San Juan Bosco

Buckle, Carlos

Universidad Autónoma de Entre Ríos

Tugnarelli, Mónica

Universidad Nacional del Nordeste

Dapozo, Gladys

Universidad Nacional de Tucumán

Luccioni, Griselda María

Universidad Nacional Arturo Jauretche

Morales, Martín

Universidad Nacional del Chaco Austral

Zachman Patricia

Universidad Nacional del Oeste

Foti, Antonio

Universidad Nacional de Cuyo

García Garino, Carlos

Universidad Nacional de Mar del Plata

Ríos, Carlos

Universidad Nacional de Quilmes

Universidad Nacional de Hurlingham

Medrano Gustavo

Universidad Nacional de San Antonio de Areco

Ramón, Hugo

Universidad Nacional de San Martín

Estayno, Marcelo

Universidad de Morón

Chapperon, Gabriela

Universidad Nacional de Luján

Fernandez, Juan Manuel

Universidad Nacional de Catamarca

Poliche Maria Valeria

Universidad Nacional de La Rioja

Molina, Miguel

Universidad Nacional de Tres de Febrero

Oliveros, Alejandro

Pontificia Universidad Católica Argentina

Grieco, Sebastián

Universidad del Salvador

Zanitti, Marcelo

Universidad del Aconcagua

Giménez, Rosa

Universidad Gastón Dachary

Ruidías, Hector Javier

Universidad Argentina de la Empresa

Feijo, Daniel José

Universidad del CEMA

Guglianone, Ariadna

Universidad Austral

Cosentino, Juan Pablo

Universidad Atlántida Argentina

Rathmann, Liliana

Universidad Católica de La Plata

Bertone, Rodolfo

Instituto Tecnológico de Buenos Aires

Bolo, Mario

Universidad Champagnat

Brachetta, Mariana

Universidad Abierta Iberoamericana

De Vincenzi, Marcelo

Universidad de Belgrano

Guerci, Alberto

Universidad Argentina John F. Kennedy

Panizzi, Marisa

Universidad Adventista del Plata

Bournissen Juan

**Universidad de la Cámara Argentina de
Comercio y Servicios**

Malbernat, Lucía

Universidad de Palermo

Alvarez Adriana

Coordinadores de Área

Agentes y Sistemas Inteligentes.

Marcelo Falappa (UNS)
Marcelo Errecalde (UNSL)
Daniel Pandolfi (UNPA)

Tecnología Informática aplicada en Educación

Zulma Cataldi (UBA-UTN)
Alejandra Zangara (UNLP)
Gustavo Gil (UNSa)

Arquitectura, Redes y Sistemas Operativos

Luis Marrone (UNLP)
Daniel Arias Figueroa (UNSa)
Orlando Micolini (UNC)

Procesamiento de señales y Sistemas de Tiempo Real

Oscar Bría (INVAP)
Fernando Tinetti (UNLP)
Nelson Rodriguez (UNSJ)

Computación Gráfica, Imágenes y Visualización

Martín Larrea (UNS)
María J. Abásolo (UNLP-UNCPBA)
Roberto Guerrero (UNSL)

Bases de Datos y Minería de Datos

Laura Lanzarini (UNLP)
Claudia Deco (UNR)
Norma Herrera (UNSL)

Ingeniería de Software

Pablo Fillottrani (UNS)
Pablo Thomas (UNLP)
Fernanda Carmona (UNdeC)

Innovación en Educación Informática

Claudia Russo (UNNOBA)
Elena Durán (UNSE)
Lucía Malbernat (UCAECE)

Innovación en Sistemas de Software

Monica Tugnarelli (UNER)
Marisa Panizzi (UK)
Guillermo Feierherd (UNTDF)
Gladys Dapozo (UNNE)

Seguridad Informática

Paula Venosa (UNLP)
Jorge Eterovic (UNLaM)
Javier Echaiz (UNS)
Antonio Castro Lechtaller (IESE)

Procesamiento Distribuido y Paralelo

Marcelo Naiouf (UNLP)
Marcela Printista (UNSL)
Javier Ballardini (UNCOMA)

Jurado Tesis Doctorado

Marcela Printista (UNSL)
Laura De Giusti (UNLP)
Silvia Castro (UNS)
Alejandra Cechich (UNCOMA)
Horacio Kuna (UNaM)

Índice

ARSO – Arquitectura, Redes y Sistemas Operativos	1
Plataformas de Automatización de Red	2
Tecnologías de Smart IoT y Aprendizaje Automático para la Solución de Problemas en el Medio Productivo	7
Estudio de la implementación del Protocolo de Internet Versión 6 en el Ámbito Regional y su Impacto en los Usuarios Finales	13
Entorno de Contenedores de Sistemas Embebidos con Conexión a Dispositivos Externos	17
Desarrollo de Dispositivos Electrónicos de Aplicación Específica a Baja Escala	22
ASI – Agentes y Sistemas Inteligentes	27
Extensiones a DEHIA en el Marco de una Arquitectura Distribuida con Microservicios	28
Aplicaciones de Machine Learning para el Uso Sustentable de Recursos Naturales	33
Generador de Proyectos de Software con Transferencia a la Comunidad	38
Deep learning para Aplicaciones Astronómicas, Visión por Computadora y Sistemas Médicos	42
Desarrollo de Técnicas de Inteligencia Computacional para el Análisis de Datos Genómicos	47
Inteligencia Computacional Aplicada a Movilidad Urbana	51
Aplicación de una Estrategia y de Técnicas de Inteligencia y Analítica de Negocio a los Sistemas de Información del Ministerio de Salud de la Provincia de San Juan	54
Optimización de la Logística de Distribución Utilizando Técnicas de la Inteligencia Artificial	59
Metaheurísticas Paralelas Aplicadas al Diseño de la Red de Sensores en Plantas Químicas	64
Gestión Sustentable del Tránsito en Ciudades Inteligentes y Sostenibles	68
Detección de la Calidad del Agua Mediante Imágenes Satelitales: Revisión Sistemática de la Literatura con Análisis Cuantitativo	72
Logística inteligente para la Recolección Dinámica de Residuos	77
Procesamiento Automático de Imágenes Médicas	82
Sistemas Conversacionales Aplicados a la Gobernanza	87
Detección Anticipada de Riesgos en la Web	92
Análisis de Calidad de Arenas de Fracturación Mediante Visión Artificial y Redes Neuronales	97
Marco de Trabajo de Rasgos Biométricos en Queiloscopía Mediante el Uso de Machine Learning	103
BDDM – Bases de Datos y Minería de Datos	108
Análisis y Aplicaciones de Internet de las Cosas y Ciudades Inteligentes Escenario de Testeo de Seguridad	109
Modelado de Variedad de Activos de Dominio en Sistemas Big Data (SBD)	113
Abordaje Metodológico de Minería de Datos Aplicada a Investigaciones Agropecuarias sobre Artrópodos	118
Estrategias Algorítmicas y Estructuras de Datos Eficientes para Búsquedas en Datos Masivos	123
Hacia la Definición de un Marco Metodológico para el Desarrollo de un Sistema de Reconocimiento Biométrico Mediante Técnicas de Machine Learning.	128

Determinación de Perfiles Socioeconómicos y Sanitarios de las Personas Atendidas en las Campañas Efectuadas como Actividades Curriculares de la Carrera de Medicina de la UNCAus en su Área de Influencia	133
Visualizando Información de Eventos Climáticos Mediante Presentaciones Interactivas Utilizando Knime Analytics	138
Propuesta de Técnicas de Validación para la Calidad de Datos Abiertos e Identificación de Patrones para Predicciones con Machine Learning	143
Aplicación de Inteligencia y Analítica de Negocios en Diferentes Contextos	148
Extracción y Explotación de Conocimiento para la Gestión en Línea de Datos en Ciencias del Mar	153
Algoritmos de Detección de Outliers	157
Chequeo de Consistencia de Procedimientos Almacenados y Triggers	161
Procesamiento Inteligente de la Información. Aplicaciones en Bioinformática, Trayectorias Vehiculares, Mantenimiento Preventivo Industrial y Sistemas Embebidos	166
Aplicaciones de Bases de Datos Espaciales y Espacio Temporales	172
Indexación y Búsquedas sobre Datos no Estructurados	176
Métodos de Series de Tiempo de Intervalos	181
Administración y Recuperación de Datos Multimedia Masivos	186
CGIV – Computación Gráfica, Imágenes y Visualización	191
Índice y Operaciones para Bases de Datos Métricas	192
Servidor Pacs Usando el Estándar Dicom - Repositorio Open Source Web DICOM en Hospital Eleazar Herrera Motta	197
Detección de Yemas Brotadas para la Estimación Temprana del Rendimiento de una Planta de Kiwi	200
Sistema de Información basado en Norma Dicom para aplicaciones oftalmológicas orientadas a Retinopatías del Prematuro	205
El Aporte de las Realidades Alternativas al Objetivo de Desarrollo Sostenible de Salud y Bienestar	210
Qgis y Bonos Verdes: Herramienta para la Clasificación de Patrones en Imágenes Geográficas Parametrizadas Según el Carbono Capturado por la Biomasa Forestal	215
IEI – Innovación en Educación en Informática	219
Aplicaciones Móviles, Realidad Aumentada y TVDi	220
Resultados Preliminares en la Formación de Docentes y Alumnos como Investigadores Científicos Iniciales	226
Modelos, Metodologías y Recursos para el Desarrollo del Pensamiento Computacional	231
Didáctica del Pensamiento Computacional y la Programación en distintos Niveles Educativos	236
Desarrollo y Evaluación de Competencias en la Ingeniería de Software en un Entorno Virtual de Aprendizaje Colaborativo	241
Diseño Participativo para Desarrollar Recursos Educativos para la Enseñanza de las Ciencias de la Computación en la Escuela Secundaria. Una Máquina de Turing en la Escuela	247
Ciencias de la Computación en la Escuela Secundaria. Percepciones Sociales y Formación Docente Inicial	252
Inteligencia Artificial: Estrategias Aplicadas a Problemas Emergentes	257
IS – Ingeniería en Software	253

Avances en la Incorporación de Estrategias Innovadoras en los Procesos de Enseñanza y de Aprendizajes de la Programación	264
Verificación Formal de Software en Sistemas de Big Data	269
Aproximación a un Método de Cálculo de Viabilidad del Despliegue de Sistemas de Software	274
Experiencia del Ciudadano: Mejorando la Interacción en Contextos Digitales	278
Proceso de Validación de Requerimientos de Software	283
Aplicaciones Móviles 3D con Realidad Virtual y Realidad Aumentada	288
Reúso de Información en Comunidades Virtuales	291
Desarrollo de Estudios Empíricos en Ingeniería de Software	296
Personalización Regional de Buenas Prácticas de Calidad de Software	301
Heurísticas para la Creación de Modelos de Evaluación Multicriterio	306
Interacción Persona-Computadora y Autismo: Interfaces que facilitan la interacción	311
Hacia la Adaptación de Scrum para Incorporar Calidad de Datos en Ámbitos del Desarrollo Ágil	316
Uso de Técnicas de Design Thinking para Mejorar los Sistemas de Producción en PyMEs	312
Creación de Software para Personas con Discapacidad Usando Design Thinking	327
Atributos de Calidad Aplicados a la Metodología Meduc_Ar para el Desarrollo de Aplicaciones en RA en Educación	332
Gobierno Digital. Evaluación del Uso de la Tecnología en el Ámbito Universitario	336
Estudio Preliminar para el Desarrollo de una Herramienta de Búsqueda en Repositorios Académicos	341
Agrupamiento de Escenarios Guiado por Objetivos	345
Utilización de los Modelos de Procesos en los Procesos de Requisitos	350
Evaluación de Calidad en Datos Abiertos. Mejora de procesos. Sistemas Resilientes en la Gobernanza Digital	355
Ingeniería de Requisitos de Software Orientada a Aspectos. Resolución de Conflictos para el Modelado Estático	360
Medición para la Evaluación Transversal de Metodologías de Gestión de Proyectos	365
Aspectos de Ingeniería de Software, Bases de Datos Relacionales, No Relacionales y como Servicios en la Nube para el Desarrollo de Sistemas de Software Híbridos	370
Análisis de Problemas Asociados al Diseño de Sistemas de IoT	377
Marco de Referencia para la Integración de Accesibilidad en Sistemas e-Learning	382
ISS – Innovación en Sistemas de Software	386
Enfoques y Tendencias en el Desarrollo de Aplicaciones Móviles con Resiliencia	387
Diseño y Desarrollo de Serious Games para la Rehabilitación de Pacientes Neurológicos Implementando VRPN para las Comunicaciones entre las Interfaces y los Dispositivos	393
Línea de Investigación Realidad Aumentada Universal Dirigida por Interacciones Procedimentales en Contextos 4.0	398
Generación de un Prototipo de Comunicación sobre Android para App Móviles Reactivas	403
Análisis de las Tecnologías 4.0 en las PyMES del Partido de Hurlingham	410
Modelos Basados en Inteligencia Artificial y Computación Ubicua para la Resolución de Problemas en Educación y Otros Dominios	414
Adecuación de un Sistema de Recuperación de Información para su Utilización en un Contexto Jurídico	418
Aplicación de Machine Learning sobre Imágenes Utilizadas en Proyectivas	423

Tecnologías de la Información Facilitadoras para la Interoperabilidad de Software en Gobierno Abierto	428
Modelo Predictivo para Evaluar el Riesgo Potencial de Existencia de Epidemia	433
Desarrollo de Middleware y aplicación Cliente para Sistema de Miniboyas Ambientales	438
Análisis e Implementación de Ciudades Inteligentes	443
Desarrollo de una Aplicación de Geofencing en Android para Monitoreo de Corredores Estudiantiles	447
Question Answering Aplicado a la Web Semántica. Predicción de la Respuesta Esperada.	452
Herramientas Informáticas de Dominio Específico para el Desarrollo de Servicios Digitales Innovadores para Comunidades Urbanas y Rurales en el Marco de Ciudades y Regiones Inteligentes	456
Coordinación de Dispositivos de IoT para Resolver Problemas Específicos de la Distribución de Energía Eléctrica	461
Modelando Ontologías con Patrones en un Ambiente Gráfico Web de Ingeniería Ontológica	467
Estrategias de Explotación de Información de Salud Pública en la Provincia del Chubut para su Uso en el Ordenamiento Territorial	472
PDP – Procesamiento Distribuido y en Paralelo	477
Inteligencia Artificial para la Multi-Clasificación de Fauna en Fotografías Automáticas Utilizadas en Investigación Científica	478
Desarrollo de un Simulador para la Evaluación de Algoritmos Clásicos y Nuevos para la Gestión de Recursos Compartidos en Sistemas Distribuidos Contemplando Exclusión Mutua	483
Diseño de Metaheurísticas Paralelas con el Paradigma Novelty Search para la Reducción de Incertidumbre en la Predicción de Fenómenos de Propagación	488
Técnicas de Modelado y Simulación en Sistemas de HPC y Salud	493
Cloud Computing, IaaS Privados y Públicos para el Análisis y Modelado de Sistemas	498
Software de Base, Modelos y Aplicaciones en Arquitecturas Multiprocesador	503
Integración de Arquitecturas Edge-Fog-Cloud en Procesamiento Distribuido. Aspectos de Eficiencia y Resiliencia	509
Migración de Aplicaciones Monolíticas a Entornos Distribuidos Serverless	514
HPC Serverless para Tratamiento de Datos Provenientes del IoT	519
Algoritmos Paralelos y Evaluación de Rendimiento en Plataformas de Cómputo de Altas Prestaciones	524
Tecnología HPC en la UNdeC como Motor de Ciencia	530
Toma de Decisiones en Sistemas de Eventos Mediante Inteligencia Computacional y Computación de Alto Desempeño	535
PSST – Procesamiento de Señales y Sistemas en Tiempo Real	540
Métricas de Rendimiento, Aplicaciones de Datos Masivos e Inteligencia Artificial en Cómputo Paralelo y Distribuido	541
Modelos Matemáticos y Aritmética Computacional	546
Redes de Sensores Inalámbricas y Simulación en Sistemas de Tiempo Real	550
Estudio del Habla de Pacientes con Enfermedad de Parkinson y Desarrollo de Aplicación Web	554
Optimización de Cálculo para la Medición de la Intensidad Acústica Compleja para Dispositivos Stand-Alone de Bajo Costo	559
Técnicas de Big Data para el Análisis de la Teoría Moderna de Optimización de Carteras de Inversión en el Marco del Mercado Argentino de Valores	562

RCCI – Redes de Cooperación Científica Internacionales	567
Formación de Recursos Humanos Orientados a Ciudades Inteligentes Sostenibles. Proyecto CAP4CITY	568
SI – Seguridad Informática	573
Consortio para la Colaboración en IDI en Temas de Cloud Computing, Big Data y Emerging Topics. Proyecto Integrador: Transformación Digital en la Incorporación de la Resiliencia como un Key Performance Indicator de Prestaciones Sociales	574
Contratos Inteligentes para Internet de las Cosas	579
Especificación Integral del Sistema OTP-Vote Orientada a su Implementación	584
Análisis de Vulnerabilidades en Videojuegos Basados en NFT y Criptomonedas	589
Protocolos de Consenso	592
Criptografía Liviana para Objetos Conectados	597
Ciberdefensa y los Usos Maliciosos de la Criptografía	600
Criptografía Liviana para Aplicar en IoT e IIoT	604
TD – Tesis Doctorales	608
Informática Forense: Métodos, Herramientas y Técnicas	609
Hybrid Networking SDN y SD-WAN: Interoperabilidad de Arquitecturas de Redes Tradicionales y Redes Definidas por Software en la Era de la Digitalización	614
Diseño y Verificación de Sistemas de Tiempo Real Heterogéneos	626
TIAE – Tecnología Informática Aplicada en Educación	636
Clasificación Automática del Grado General de Disfonía	637
Gobernanza de Datos en los Procesos de Negocio para las Instituciones de Educación Superior	647
Anexando Acciones en el Aprendizaje Móvil	652
Modelo de Evaluación de Software Educativo Libre Basado en un Método Sistemático	657
Aplicación de Analítica de Datos en Espacios de Educación Superior	662
Construcción de Conocimientos en la Universidad a Partir del Diseño e Implementación de Escenarios Tecno-Pedagógicos y Materiales Educativos Digitales.	667
Modelos de Desarrollo de Serious Games. Las Analíticas de Aprendizaje e Inteligencia Artificial	673
Modelo de Diseño de Aprendizaje para la Enseñanza de la Física en una Modalidad e-Learning Usando Laboratorios Virtuales	678
Realidad Aumentada en Escenarios Educativos Científicos	682
Desarrollo de Aplicaciones para Colaboración en e-Learning	687
Construcción de una Herramienta de Soporte para la Creación de Diagramas de Clases Destinados a Estudiantes con Disminución Visual.	692
Recomendaciones Utilizando Inteligencia Artificial a Partir de las Métricas de Evaluación Obtenidas de las Analíticas de Aprendizaje Aplicadas a Serious Games	697
Investigaciones y Experiencias en el Área de Interacción Persona Ordenador y Educación	702
Realidad Aumentada en Diálogos de Estudiantes Colaborativos	707
Adaptaciones en Aprendizaje Colaborativo Soportado por Computadora	711
Nuevos Criterios para Formación de Grupos Colaborativos en Collab	716
Sistema Informático de Predicción de Deserción Estudiantil Universitaria Basado en un Modelo de Indicadores de Regulación del Aprendizaje, en Entornos Educativos Mediados por TIC	720
Diseño de Sistemas Recomendadores en el Ámbito de las Bibliotecas Universitarias	725
Aproximación al Diseño de Aprendizaje Aplicado a un Curso de e-Learning	729

Estrechando el Contacto Entre Universidades y Estudiantes: Comunicación Ante Posibles Casos de Abandono, Propuestas para la Inscripción. Línea de Investigación y Desarrollo del CIDIA.	734
Educación a Distancia e Innovación Tecnológica	739
Entornos Virtuales y Herramientas Digitales en Escenarios Educativos Híbridos	744
Herramientas de Software para Soporte en la Enseñanza y Aprendizaje de la Arquitectura x86	751
Evolución Temática de Publicaciones en Español. Su Aplicación en el Campo de la Tecnología y la Educación	756
Dispositivos Móviles en Colegios de Nivel Secundario	760
Herramientas Digitales para Educación. Análisis de su Accesibilidad	765
Objetos de Aprendizaje: Diseño y Construcción Basado en Criterios de Calidad	771
Explorando con Realidad Virtual Interactiva	776

Autores

A

Abasolo, María: 220, 398
Aciti, Claudio: 662
Acosta, Denise: 472
Acosta, Derin: 414
Adagio, Matias: 17
Agostini, Federico: 483
Aguilar, Norma: 197
Aguilera, C: 54
Aguirre, Agustín: 776
Aguirre, Emmanuel: 133
Aguirre, Verónica: 387
Ahmad, Tamara: 739
Aidelman, Y: 42
Alaniz, Mariano: 148
Alba, M. D: 192
Albornoz, Claudia: 311
Alcarado, Y: 210
Aldarete, Claudia: 443
Alfonso, Hugo: 59
Almarcha, Matias: 257
Álvarez, Margarita: 414
Alvez, Carlos: 118, 128
Amadeo, Paola: 765
Amet, Leonardo Javier: 33
Antelo, Micaela: 17
Antonelli, Leandro: 283
Aranda, Gabriela: 291
Arce, Camila: 716
Arcidiáceno, José: 28
Argañaraz, Julián: 393
Arias Figueroa, Daniel: 13
Arias, Daniel: 514
Arias, Silvia: 257
Armas, Adrián: 597
Armas, Alfredo: 157
Aroyuelo, J: 192
Arsaule, Ariel: 161
Artola, Verónica: 702
Ascheri María: 657, 760
Asteasuain, Fernando: 269
Astudillo, Gustavo: 264

Atencio, Hernán: 514
Ayala, Ludmila: 760
Azar, Paola: 176

B

Bacigalupe, María: 720
Balaguer, Federico: 200
Baldassarri, Sandra: 702
Balladini, Javier: 503, 541
Banceloff, Claudia: 776
Banchemero, Santiago: 123
Bancolini, Alessandro: 266
Baquela, Enrique: 51
Barbero, Antonella: 264
Barrera, Elizabeth: 597
Bast, Silvia: 264
Barrivera, A: 166
Basti, Silvia: 584
Basum, Matias: 7
Battaglia, Nicolás: 241
Bazán, Patricia: 28
Becerra, Martín: 398
Bemón, Mario: 377, 584
Bender, Cristian: 236
Benitez, Pedro: 118
Berecochea, José: 133
Bermúdez, Carlos: 59, 64
Bernaldo de Quiroz, Franco: 336
Bertolo, Regina: 97
Bertone, Rodolfo: 38, 274, 410
Bianchini, Germán: 266, 488
Bianco, Santiago: 756
Bilbao, M: 77
Bizaid, Gustavo: 205
Blanco, Gabriel: 350
Blanco, Pedro: 33
Boari, Sandra: 252
Bolatti, Diego: 109
Bond, Román: 493, 498
Bono, Sofía: 316
Borachia, Marcos: 355
Botta, Cristian: 7

Bouciquez, María: 220
Brandan, F: 87
Bria, Oscar: 546
Britos, Luis: 186
Britoz, José: 257
Buccella, Agustina: 113
Buckle, Carlos: 153, 472
Burdiso, Sergio: 92
Bustamante, Gimena: 478
Bustillo, Mario: 682

C

Cabrera, Sergio: 725
Caccaviello, Diego: 438
Cáceres, Nélica: 341
Cáceres, Raquel: 332
Cáceres, Rubén: 133
Cafferri, Yanina: 215
Cagnina, Leticia: 92
Caldoira, Susana: 269
Camiletti, Paulo: 760
Company, Ana: 231
Cantella, A: 77
Capdevila, Juan: 148
Cappolletti, Marcelo: 7
Carbadillo, Jessica: 47
Cardozo, Claudia: 278
Cardozo, Franco: 197
Carmeletto, G: 166
Carmuccio, Esteban: 17
Carnero, Mercedes: 64
Carnuccio, Esteban: 447
Caro, Angélica: 316
Carrizo, Claudio: 316
Carro, Victoria: 257
Carrona, Fernanda: 771
Carruthers, Juan: 296, 301
Caruso, Martín: 143
Carvalho, Gastón: 51
Casali, Ana: 236
Casanova, Carlos: 172, 535
Cassola, Guillermo: 514
Castillo, Gonzalo: 316
Castro, Artemio: 771
Castro, Iván: 382

Castro, Leandro: 264
Catania, Carlos: 562
Cavaliere, Lucas: 291
Caymes, Paola: 226
Ceballos, Darío: 153
Cecchi, Laura: 467
Cechich, Alejandro: 113
Cerrada, María: 166
Challiol, Cecilia: 321, 327
Chatterjee, Paraj: 103
Chávez, Edgar: 192
Chávez, Susana: 403
Chazarretta, Facundo: 7
Checatelli, Diego: 148
Chichizola, Franco: 503, 524
Chirino, Pamela: 226
Chuck, Oscar: 97
Cicerchia, Benjamín: 739
Cipriano, Marcelo: 579, 600, 604
Collazos, César: 220
Colombani, Marcelo: 751
Colombes, Mariela: 274
Colussi, Natalia: 236
Constanzo, Manuel: 503, 509, 524
Corbalán, Leonardo: 387
Cornersi, Sandro: 257
Correa, Martín: 38
Cortese, Rocío: 316
Cortez, Marcos: 252
Costaguta, Rosanna: 707, 711, 716
Coyazos, César: 336
Cristaldo, Patricia: 365
Cuadrado, María: 157
Cuadros, Patricia: 667

D

Dal Bianco, P: 42
Daniele, Marcela: 161
Dantiacq, Alejandro: 2
Dápolo, Sebastián: 288
Dapozo, Gladys: 231
Dapozo, Gladys: 301
Dasso, Arístides: 306, 433
David, María: 200

Dávila, Laura: 257
De Battista, Anabella: 176, 365
De Miranda, Norcelo: 72
Deco, Claudia: 236
Decoud, Carla: 739
Degaran, Gustavo: 200
Del Bono, Dana: 257
Del Castillo, Florencia: 472
Del Do, M: 68
Delgado, César: 447
Delia, Lisandro: 387
Delvechio, Tomás: 123
Denon, Nicole: 7
Di Genaro, M: 192
Díaz, Beatríz: 118
Díaz, Canin: 226
Díaz, Fátima: 414
Díaz, Javier: 765
Dielschneider, Juliana: 59
Dieser, Paula: 744
Dodaro, Luciano: 428
Dogliotti, Mariano: 443
Dolz, Daniel: 247
Domínguez, Facundo: 438
Doña, Pablo: 559
Doom, Jorge: 345
Dorzán, María: 172
Doti, Santiago: 37
Drazic, Leandro: 148
Droz, Mariano: 118
Duarte, David: 498
Durán, Elena: 414
Durán, Elena: 678

E

Encinas, Diego: 7, 493, 498
Eguren, Santiago: 562
Errecalde, M: 87, 92
Escudero, C: 42
Espeche, Fabián: 678
Espíndola, María: 231
Esponda, Silvia: 355
Estrebou, C: 166
Etchart, Graciela: 128
Eterovic, Jorge: 579

Evin, Diego: 205

F

Faccini, Higinio: 2
Falappa, Marcelo: 751
Fara, Laura: 776
Federico, Cristina: 288
Fernández, Gladys: 749
Fernández, J: 210
Ferraro, María: 301
Ferrer, Liliana: 682
Figueroa, Karina: 192
Figueroa, Liliana: 609
Figueroa, Patricia: 530
Fillottrani, Pablo: 467
Florencia, María: 103
Flores, Andrés: 291
Flores, Natalia: 332
Flores, Sebastián: 377
Flores, Sergio: 403
Fogliro, Alejandro: 157
Fourcade, Alejandro: 22
Fрати, Emmanuel: 524, 530, 541
Fritz, Estela: 657
Frutos, Mariana: 161
Funes, Ana: 306, 433
Funes, Darío: 92

G

Gardiala, Joaquín: 161
Gaetán, Gabriela: 274
Gaiani, Magalí: 734
Galdamez, Mariela: 226
Galvez, María: 332, 341
Gamarra, Álvaro: 13
Gamboa, Pablo: 269
Gamer, R: 42
García, Edith: 600, 604
García, Leandro: 711
García, Mario: 82, 637
García, Rubén: 530
García, Sergio: 418
García, Pablo: 657
Garciaarena, José: 92

Gargano, Cecilia: 418
Gaudini, Adriana: 524
Gavilanes, Wilma: 220
Gel, Matias: 478
Gerzel, Stella: 133, 483
Giacomantone, Javier: 546
Giménez, Cristian: 467
Gira Díaz, Facundo: 776
Giuliano, Mónica: 200, 554, 734
Giusti, Laura: 503, 509, 524
Giusto, Luana: 33
Giustti, Armando: 220, 503, 509, 524, 541
Glagiardi, Edilma: 172
Godoy, María: 301
Gómez, Adriana: 148
Gómez, Martín: 514
Gómez, Mauro: 493
Gómez, Pablo: 519
González, Agutín: 123
González, Alejandro: 744
González, Federico: 478
González, Gabriela: 414
González, Rodrigo: 393, 584
Gorga, Gladys: 702, 744
Grafigna, Carlos: 559
Gramajo, Sergio: 109
Gramática, Martín: 82
Granata, Nicolás: 269
Greiner, Cristina: 231, 301
Grosso, A: 192
Guccione, Leonel: 673, 697
Guerrero, R: 210
Guevara, Diego: 148
Guidet, Sebastián: 559

H

Hadad, Alejandro: 205
Haderne, Marisa: 51
Hang, Shao Feng: 269
Harrari, Ivana: 765
Haspervé, W: 166
Hernández, José: 64
Herrera, Norma: 176
Herrera, Susana: 609
Hidalgo, Fabián: 2

Hirschfeld: Gisela: 257

I

Ierache, Jorge: 398
Iglesias, Luciano: 702
Illescas, Gustavo: 662
Inchausti, Pablo Ezequiel: 33
Irrazabal, Emanuel: 296, 301
Isturan, Romina: 720
Izaguirre, Evelyn: 220

J

Jofré, M: 192
Jofré, N: 210
Juarez, Carlos: 678
Juarez, Gastón: 414

K

Kaplan, Gladys: 345, 350
Kasian, Fernando: 186
Klenzi, Raúl: 138
Kuna, Horacio: 72

L

Laceiras, Silvia: 72
Lanciotti, Julieta: 503
Lanzarini, L: 166
Lanzarini, L: 42
Lanzarini, Laura: 756
Lanzillota, Franco: 673
Lasagna, Valeria: 720
Lavallén, Pablo: 123
Lebetti, Joaquín: 519
Lewis, Mirtha: 153
Lezcano, Andrea: 296, 301
Literas, Alejandra: 28
Lobos, Martín: 264
Lomero, Laura: 579
Lomoro, Jorge: 241
López, Diego: 143
López, P: 166
Lopresti, Mariela: 535

Loyola, Martín: 92
Lucero, Sandra: 291
Ludueña, Verónica: 186, 192
Luna, Adriana: 97
Lund, María: 148
Lupi, Augusto: 438
Luzuriaga, Juan: 113

M

Maccallini, Lucas: 493
Maffei, Fabián: 241
Magnini, Giulana: 257
Maldonado, Charles: 241
Mallea, Adriana: 181
Mallicundo, Felipe: 341
Manassero, Martín: 148
Marrero, Luciano: 370
Martíarez, Nestor: 257
Martín, Adriana: 278, 403
Martín, Sofía: 776
Martínez, Ana: 33
Martínez, Cecilia: 181
Martínez, David: 133, 483
Martínez, Enrique: 197
Martínez, Gabriel: 197
Martínez, Luis: 707
Martínez, Nadira: 291
Martínez, Rodolfo: 113
Martínez, Roxana: 428
Martínez, Sergio: 341
Marzalu, Rafaela: 113
Masanet, María: 138
Massa, Stella: 673, 697
Massej, Florencia: 410
Mazzanti, Renato: 153
Medina, Santiago: 550
Medina, Yanina: 231, 296
Melani, Martín: 609
Méndez, Nicolás: 498
Menini, Ángeles: 687
Menini, María: 707
Mente, Augusto: 336
Mercado, V: 68
Migani, Silvina: 148
Mindlin, I: 42

Minetti, Gabriela: 59, 64
Minetti, Yamila: 264
Mir, Sebastián: 336
Miranda, Ernesto: 128
Miranda, Franco: 186
Miranda, Natalia: 535
Molina, D: 68
Molina, María: 148
Monjelat, Natalia: 236
Montaldo, Abril: 478
Montejano, Germán: 311, 584, 589
Montes De Oca, Federico: 493, 498
Montezanti, Diego: 7
Monzón, Francisco: 123
Morales, Martín: 7, 493
Morán, Marina: 541
Moreno, Juan: 692
Moreno, Leandro: 274
Morila, Ezequiel: 257
Moro, Lucrecia: 673
Moyano, Gladys: 257
Moyano, Marcelo: 113
Muñoz, Horacio: 456
Muñoz, Martín: 138
Muñoz, Rocío: 355
Murazzo, María: 514, 519
Murillo, Natalia: 200

N

Naiouf, Marcelo: 220
Neil, Carlos: 241
Nieto, Facundo: 257
Nieves, Guillermo: 33
Núñez, Enrique: 97
Núñez, Gustavo: 153

O

Ojeda, Celeste: 296
Ojeda, Juan: 418
Oliva, Facundo: 241
Olivar, Matias: 423
Olivares, Diego: 186

Olivera, Ana: 47, 51
Olivera, Lucas: 7
Olsowy, Verena: 370, 387
Ontiveros, Patricia: 226
Orozco, M: 77
Ortíz, Claudia: 327, 321
Ortíz, Felipe: 410
Ortíz, Sara: 403
Osorio, Alejandra: 765
Osorio, Jorge: 7
Osycka, Liam: 113
Oyarzun, Matias: 452

P

Pacífico, Cristian: 725
Padilla, Jonatan: 197
Paez, Francisco: 626
Palacios, Cristian: 197
Palma, Marcelo: 370
Palmero, Pablo: 172
Pandolfi, D: 77, 87
Pandolfi, E: 68
Panizzi, Marisa: 274, 370,410,423,729
Paredes, Rodrigo: 192
Parkinson, Christian: 143
Parra, Gerardo, 247
Parra, Lorena: 514, 519
Pascal, Andrés: 176
Pasini, Ariel: 355, 370
Patera, Nahuel: 269
Pazo, María: 600
Peniego, Juan: 509
Peña, M: 166
Peralta, Lorena: 692
Peralta, Mario: 316, 382
Perdomo, Luciano: 472
Pereyra, N: 42
Pérez, D: 77
Pérez, Gustavo: 418
Pérez, Ignacio: 478
Pérez, Santiago: 2
Pérez, Silvia: 200
Pérez, Silvia: 554
Perriego, Juan: 550

Pesado, Patricia: 220, 355, 3770, 387,
744
Petinari, Joaquín: 410
Petris, Raquel: 231
Petroff, Marcelo: 456
Píccoli, Fabiana: 519, 535
Pinazzola, Luis: 609
Pincioli, Fernando: 360
Pinto, Sonia: 148
Pizarro, Rubén: 760
Pizarro, Rubén: 461
Pollo, Cattaneo: 103
Pollo, Cattaneo: 103
Portugal, Juan: 393
Printista, Marcela: 186
Puppo, Florencia: 220
Purichelli, Fernando: 410

Q

Quieroga, Ariel: 197
Quieroga, F: 42
Quintana, Fabio: 418
Quiñonez, Tomás: 467
Quiroga, Rubén: 197
Quiroz, Daniel: 37

R

Ramón, Hugo: 336
Ramos, Juan: 118
Ramos, L: 68
Rasjido, J: 77
Ravlic, Ignacio: 200
Retamar, Soledad: 365
Reyes, Nora: 186, 192
Riesco, Daniel: 377
Rios, G: 42
Rísola, Esteban: 120
Riva, Alberto: 771
Rivarola, Federico: 428
Rivera, Paula: 530
Rivero, Julieta: 720
Riveros, Alejandro: 148
Roberti, Bruno: 2
Rodrigo, Alexis: 403

Rodríguez, Ayelén: 410
Rodríguez, Carlos: 97
Rodríguez, G: 210
Rodríguez, Guillermo: 662
Rodríguez, Jorge: 247, 252
Rodríguez, Nelson: 514, 519
Rodríguez, Rocío: 443, 447
Rodríguez, Sebastián: 503
Roger, Sandra: 452
Roguo, Patricia: 186
Rojas, Matias: 47
Rojas, Nayiby: 143
Romagnaro, N: 54
Romero, Fernando: 550
Romero, Mara: 530
Ronchetti, F: 42
Ronchetti, Franco: 220
Rosado, Magdalena: 220
Rosatto, Daniel: 493
Rosestein, Javier: 393
Rosete, A: 42
Rossi, Carlos: 197
Rotella, Carina: 226
Rozas, Claudia: 541
Ruano, Dario: 176
Rucci, E: 42
Rucci, Enzo: 503, 530
Ruiz, Hector: 456
Ruiz, José: 751
Ruiz, Silvia: 128
Russo, Claudia: 739
Ryckeber, Hugo: 418

S

Sabelli, Agustín: 103
Saisar, Victoria: 418
Salas, Andrea: 181, 370
Salazar, Gustavo: 614
Salazar, Mesia: 702
Salazar, Nevelin: 414, 711
Saldaño, Viviana: 278
Saldarini, Javier: 316
Saldivia, Cristian: 734
Salgado, Carlos: 316
Salgado, Carlos: 382

Salinas, Héctor: 393
Salinas, Sergio: 226
Salto, Carolina: 59, 64
Salvatelli, Adrián: 205
Salvatore, Juan: 7
Samat, Pablo: 186
Sambrana, Iván: 296
Sánchez, Alberto: 316
Sánchez, Ernesto: 13
Santana, Pablo: 687
Santana, Pablo: 716
Santana, Sonia: 283
Santander, Candela: 345
Santirelli, Paola: 336
Santos, Graciela: 220
Sanz, Cecilia: 220, 702, 744
Sarobe, Mónica: 739
Sarra, José: 771
Sattolo, Iris: 423
Sattolo, Iris: 729
Scappini, Reinaldo: 109
Schab, Esteban: 535
Schiavoni, Alejandra: 765
Schiffino, Cristian: 410
Schroeder, Romina: 291
Scutari, Paola: 488
Segovia, Darío: 461
Sigampa, Elvio: 197
Signorini, Lucas: 257
Silva, Yelmo: 220
Silvestrini, Pía: 662
Sosa, Hernán: 274
Spósitto, Osvaldo: 418
Stanzial, Doménico: 559
Stragga, Juan: 488
Suppi, Reno: 541

T

Tagarelli, Sandra: 226
Taramilla, María: 172
Taruschio, Marcelo: 38
Tejada, Jorge: 530
Testa, Oscar: 461
Texier, José: 530, 771
Teze, Juan: 725

Thomas, Pablo: 283, 370, 387
Thomson, Horacio: 92
Tinetti, Ferrando: 550
Tinetti, Rodrigo: 503, 509
Tolaba, Ana: 332, 288
Tolosa, Gabriel: 123
Torres, Juan: 335
Torres, Luis: 579, 604
Torres, M: 68
Tost, Carolina: 109
Tucci, Trinidad: 257
Tugnarelli, Mónica: 739

U

Únzuga, Silvina: 414
Ureta, Laura: 667

V

Valiente, Waldo: 17
Vallejo, Selene: 133
Varas, V: 68
Vardano, Sergio: 682
Vargas, Alejandro: 215
Vargas, Javier: 59
Vargas, Rocío: 143
Vasallo, Alejandra: 148
Vega, Leonor: 257
Vera, Cristina: 148
Vera, Pablo: 447
Viale, Pamela: 236
Viavattene, Hernán: 447
Vidal, Pablo: 47, 51
Viera: Leonel: 478
Vignoli, Adolfo: 257
Villagarvia, Horacio: 503
Villagra, A: 68, 77, 87
Villagra, Silvia: 274
Villán, Marco: 589
Villarrubia, Laura: 341
Villegas, Carolina: 113
Villegas, Paula: 92
Vincenzi, Marcelo: 241
Vivas, Luis: 546
Volker, Mariano: 17

Volman, Mariel: 47, 51

W

Wehbe, Ricardo: 597
Wernik, Diego: 428
Wernli, Analía: 257

Y

Yommi, Alejandra: 200

Z

Zambrano, G: 166
Zambrano, Jusmeidy: 530
Zamudio, Eduardo: 72
Zapata, Luz: 118
Zaradnik, Ignacio: 438
Zárate, Ignacio: 593
Zárate, Marcos: 153
Zeballo, Eduardo: 257
Zeligueta, Laura: 360
Zengara, Alejandra: 756
Zoratto, Valeria: 291
Zorzán, Fabio: 161

ARSO – Arquitectura, Redes y Sistemas Operativos

Plataformas de Automatización de Red

Santiago Pérez, Higinio Facchini, Alejandro Dantiacq, Bruno Roberti, Fabián Hidalgo
CeReCoN (Centro de Investigación y Desarrollo en Computación y Neuroingeniería)
Facultad Regional Mendoza, Universidad Tecnológica Nacional
Rodríguez 273, Mendoza, Argentina - 0261-5244576
(*santiagocp,higiniofac,alejandrod*)@frm.utn.edu.ar

Resumen

En la línea de tiempo dada por las apariciones del open software, del open hardware, y del open networking, se favorece la aspiración concreta de contribuir a la automatización de las redes. Diferentes empresas, corporaciones, CSPs, y especialistas han puesto énfasis en la dificultad de gestión de las redes modernas, que introducen, luego de 50 años de interconexión TCP/IP, novedades tecnológicas de alto impacto, como el cloud computing, la movilidad, los nuevos perfiles de tráfico, la NFV, el IoT, el Bid Data, entre otras.

La automatización de redes es una metodología en la que usando software se configura, se aprovisiona, se administra y se prueban automáticamente los dispositivos de la red. La mayoría de los grandes operadores de comunicaciones trabajan y tienden, en este sentido, a redes verdaderamente autónomas, que a su vez, eventualmente, requerirán el uso intensivo de la Inteligencia Artificial (IA).

Este trabajo propone un estudio sobre el estado del arte en la automatización de red, planteando una visión general desde lo conceptual y, luego, avanzando sobre las tendencias vistas en los ambientes IT de las grandes empresas y operadores de comunicaciones más involucradas en esta tecnología y, finalizando, aspira a una discusión de las herramientas de automatización de red más reconocidas.

El personal principal relacionado a esta línea de investigación es docente de grado y posgrado de la Universidad Tecnológica Nacional, Regional Mendoza, y tesis de

posgrado en el Doctorado en Ingeniería (mención Computación), y de la Especialización en Redes de Datos de la misma institución.

Palabras clave: Automatización de Red, Herramientas de Automatización, Redes de Datos.

Introducción

Se denomina automatización, en general, al acto y la consecuencia de automatizar. Este verbo, por su parte, alude a hacer que determinadas acciones se vuelvan automáticas, es decir, que se desarrollen por sí solas y sin la participación directa de un individuo. En otras palabras, automatizar un proceso es conseguir que, aplicando un mecanismo de realimentación funcione sin intervención humana. Sin dudas estos términos tienen su origen en la industria manufacturera, mucho antes que existieran las redes, tal como las conocemos ahora.

Sin embargo, aunque alguna definición de automatización de red puede conservar la idea general planteada en el párrafo anterior, para la automatización de las redes existen numerosas definiciones, que varían en el alcance y especificidad del concepto de lo que involucra la automatización de red. De hecho, algunos vendedores de software las denominan herramientas de automatización, mientras que otros las denominan herramientas de administración de configuración, aparentemente para la misma función. En [1] se define que “La automatización de la red es el proceso de automatización de la configuración, administración y operaciones de una red

informática. Es un término amplio que incluye una serie de herramientas, tecnologías y metodologías utilizadas para automatizar los procesos de red". Mientras que en [2] la definición no es mucho más clara y la define como una "frase utilizada para describir las metodologías, procesos y tecnologías necesarias para ayudar a las grandes organizaciones y empresas a automatizar la configuración y administración de dispositivos de red, incluidos concentradores, enrutadores y conmutadores".

De tal forma, entonces, que según estas definiciones, se puede entender que la automatización de la red es el proceso de automatizar diferentes tareas que normalmente realizan los administradores de la red. Es decir, cualquier proceso que sea autónomo, que reduce y potencialmente elimina la necesidad de intervención humana. Mientras, y a pesar de ello, es válido reconocer que recientemente se ha producido un aumento en la cantidad de herramientas que asisten en la automatización de redes, algunas de las cuales son más reconocidas que otras en este ambiente, y que han marcado un cambio en la forma en que los administradores las construyen y administran. Y si bien las herramientas de automatización de servidores y aplicaciones han existido durante algún tiempo, pocas herramientas de este tipo han estado disponibles para las redes.

Por lo expuesto, resulta de interés precisar el alcance práctico de la definición, plantear un análisis comparativo de las diferentes y principales características de las herramientas, que en estos tiempos son reconocidas como herramientas de automatización de red, y/o herramientas de administración de configuración de red, entre ellas a Ansible, Chef, Puppet y SaltStack

Estado del Arte

La automatización de redes es una metodología en la cual el software configura, aprovisiona, administra y prueba automáticamente los dispositivos de red. Lo

utilizan las empresas y los proveedores de servicios para mejorar la eficiencia y reducir los errores humanos y los gastos operativos.

Un informe reciente de MIT Technology Review [3], producido en asociación con Ericsson [4], define la automatización de redes como "la eliminación de tareas manuales repetibles y su reemplazo por tareas programadas automatizadas con el uso de software". Los ejemplos incluyen la configuración de servidores, la programación del mantenimiento y la adición o eliminación de servicios. La automatización de la red que va más allá de un único servidor o servicio a la configuración de varias funciones de red virtualizadas, a menudo, requiere cierta orquestación para administrar los flujos de trabajo en la red. Un buen ejemplo de esto es la prueba exitosa de Vodafone [5] de la automatización completa de sus servicios de conectividad de transporte de tráfico.

Por ello, es necesario introducir nuevos conceptos y definiciones relacionadas a la automatización de red. Entre ellas:

- Circuito cerrado: Blue Planet [6], una división de Ciena, ofrece una buena definición de circuito cerrado: "un ciclo continuo y repetitivo de comunicaciones entre la infraestructura de red y los elementos de software, que incluyen análisis, políticas y orquestación, para permitir capacidades de optimización automática".
- Capacidades de optimización automática: la optimización automática lleva el circuito cerrado un poco más allá, aprovechando los procesos de circuito cerrado, para ajustar automáticamente los parámetros y configuraciones, para hacer un uso óptimo de los recursos restringidos, como recursos informáticos, radios, instalaciones de transporte y acceso, y energía.
- Red autónoma: las redes totalmente autónomas aún no existen, pero los miembros de TM Forum las definen como "proporcionar el ciclo de vida del servicio, a pedido, con una intervención humana mínima o nula".

La idea es que las redes autónomas configuren, supervisen, mantengan y reparen de forma independiente, proporcionando un conjunto de servicios de red y TIC totalmente automatizado, sin esperas, sin contacto y sin problemas para empresas en muchos sectores verticales de la industria y los consumidores. La mayoría de los grandes operadores CSP (Communications Service Provider) creen que las redes autónomas eventualmente requerirán Inteligencia Artificial (IA).

Estado de Avance

La automatización ya está creando eficiencias para los CSP. China Mobile [7], por ejemplo, ha podido mejorar el rendimiento en las redes de múltiples operadores en un 14,5%, a través de la optimización iterativa que permite la mejora continua en sus algoritmos basados en IA. Los datos resultantes han permitido a China Mobile seleccionar automáticamente velocidades de Internet más rápidas, para los consumidores que usan el mismo tipo de dispositivo. China Mobile demostró este beneficio en la prueba de concepto AIOps Catalyst (Artificial Intelligence for IT Operations Catalyst) en mayo de 2019 [8].

China Mobile y China Telecom [9] también redujeron el consumo de energía en un 10% durante el Catalyst. La reducción se logró cambiando automáticamente a los usuarios a bandas más bajas del espectro, en situaciones de poco tráfico y apagando las bandas más altas hasta que fuera necesario. Apagar las bandas más altas no solo ahorra costos de energía, sino que permite a los operadores cubrir un área más amplia con menos radios activas.

Durante un seminario web del TM Forum, del Global Architecture Forum [10], sobre redes autónomas, Orange [11] proporcionó los resultados de un estudio interno realizado para determinar el potencial de creación de valor y ganancias de eficiencia, de las redes autónomas, en las áreas de operaciones y mantenimiento, energía, utilización de recursos y aplicaciones.

La automatización tiene sus propias características: 1) La herramienta de automatización puede trabajar las 24 horas del día, sin interrupciones, lo que resulta en una mayor eficiencia; 2) La automatización permite la recopilación de grandes cantidades de datos, que pueden analizarse rápidamente, para proporcionar información que pueda ayudar a guiar un evento o proceso; y 3) En ciertas circunstancias, una herramienta de automatización de red inteligente pueden alterar su comportamiento, con un cierto objetivo.

Estas grandes empresas han aplicado la automatización de red considerando tres beneficios principales.

- Eficiencia mejorada: Al automatizar las funciones en los dispositivos de red, los humanos ya no tienen que realizar tareas que requieren mucho tiempo.
- Probabilidad reducida de error humano: Las tareas manuales son propensas a errores humanos, y cuando la estabilidad de una red de empresa o proveedor de servicios está en juego, las consecuencias de un error de cálculo o una entrada incorrecta pueden ser significativas. Configurar una tarea para la automatización significa que solo debe ingresarse correctamente una vez.
- Menores gastos operacionales: Este beneficio viene como resultado de los dos elementos anteriores. Al eliminar ciertas tareas manuales relacionadas con el aprovisionamiento de dispositivos de red y la administración de redes, las empresas pueden operar con mayor velocidad y agilidad. Por ejemplo, el aprovisionamiento automatizado puede evitar que un ingeniero de redes tenga que viajar a una nueva sucursal, para establecer la conectividad de la red, lo que permite a los empleados de ese sitio trabajar más rápido.

Una clasificación de las herramientas de automatización tiene en cuenta el origen de las mismas en: 1) Herramientas de automatización de infraestructura, 2) Herramientas de automatización de propósito específico, y 3) Herramientas de LAN Definida por Software.

La primera categoría son las herramientas que se crearon principalmente para la automatización de aplicaciones y servidores, pero se han expandido para incluir la red. Es posible que muchas TI empresariales ya tengan algunas de estas herramientas en uso, para fines de infraestructura y DevOps. El beneficio de estas herramientas es que ya son internas y el personal de TI puede estar bien versado sobre cómo usarlas. Por lo tanto, agregar procesos de automatización de red a una herramienta existente no sería muy difícil desde una perspectiva de costo, implementación y curva de aprendizaje. Dicho esto, algunas de estas herramientas de automatización de infraestructura son ligeras en cuanto a características de automatización de red y compatibilidad con diversos hardware y software de red. frecuentes en la configuración de LAN. Es decir, estas herramientas están en la parte superior, y se encuentran totalmente integradas, dado que tienen todo tipo de funciones, como monitoreo de rendimiento, análisis de tráfico y ancho de banda, configuración y gestión de cambios, puerto y conmutador de monitoreo y seguimiento de usuarios finales, monitoreo de rendimiento de WAN y gestión de direcciones IP. Estas herramientas a menudo están destinadas a reemplazar la mayoría de las otras herramientas y ofrecen una solución integral.

Otro tipo de plataforma de automatización de red son las herramientas diseñadas específicamente para la red. Estas herramientas ofrecen las características y funcionalidades más específicas de la red. La advertencia, por supuesto, es que es otra herramienta única para este propósito específico. Las tiendas de TI con exceso de herramientas pueden querer buscar en otro lado. Es decir, que están en el otro extremo del espectro. Son herramientas mucho más específicas, como las herramientas para la gestión de la configuración. Si bien estas herramientas solo manejarán la configuración del dispositivo, siguen siendo muy útiles y a menudo son accesorios para cumplir con varios estándares regulatorios. En el medio hay

muchas herramientas diferentes que ofrecen diferentes niveles de automatización. En términos generales, las herramientas con más funciones tenderán a ser más caras, aunque este no es siempre el caso. Pero no debe caerse en la tentación de dejar que el precio influya al evaluar el valor de una herramienta. Una herramienta menos costosa podría ser una mejor opción según las necesidades específicas. Por ejemplo, si se tiene alguna herramienta de monitoreo, probablemente no se necesite una herramienta de automatización que incluya funciones de monitoreo. Después de todo, es probable que se haya pasado mucho tiempo configurando herramientas de monitoreo y no se quiera desperdiciar esfuerzos.

Por último, las plataformas definidas por software crean una superposición de software en el hardware de LAN. Esta superposición permite a los administradores administrar y orquestar centralmente las configuraciones desde un único panel de administración, enmascarando los comandos de configuración subyacentes que están ocurriendo dentro de la capa subyacente. Esta opción es ideal para simplificar entornos de múltiples proveedores que requieren cambios.

Objetivos

Objetivo General:

El presente trabajo tiene como objetivo plantear un análisis comparativo de las diferentes y principales características de las herramientas, que en estos tiempos son reconocidas como herramientas de automatización de red, y/o herramientas de administración de configuración de red.

Objetivos Específicos:

Para el presente proyecto se buscan los siguientes objetivos:

- Determinar el impacto de avance de estas tecnologías, y la diversidad de modelos arquitectónicos disponibles.

- Determinar las características a detalle de las herramientas de automatización, y de los escenarios para su experimentación.
- Determinar las configuraciones óptimas SDN y de red tradicionales, y sus ámbitos de aplicación, para cada uno de los escenarios de experimentación de la automatización.
- Definir las necesidades actuales locales a fin de iniciar desarrollos y experiencia, para alcanzar una masa crítica de conocimientos y poder interactuar, con otros investigadores y profesionales.

Metodología

Para el desarrollo del proyecto se ha previsto las siguientes tareas:

Tarea 1: Recopilar de información y estudio de diferentes herramientas de automatización de red disponibles.

Tarea 2: Compilar trabajos de investigación sobre la temática, determinando resultados experimentales, herramientas de simulación, tráfico utilizados, métricas medidas, resultados y conclusiones.

Tarea 3: Establecer mecanismos comparativos de los trabajos de investigación compilados, usando cuadros, índices, ponderaciones, etc.

Tarea 4: Definir escenarios de automatización que faciliten el contraste entre diferentes tipos de tecnologías de automatización.

Tarea 5: Construir tablas comparativas de las métricas para cada escenario. Se construirán tablas y graficas comparativas de las prestaciones o métricas para cada caso, según las combinaciones que permitan los diversos escenarios.

Formación de Recursos Humanos

Mediante el presente proyecto se busca la formación de Recursos Humanos con el fin de:

- Aglutinar RRHH calificados en torno a estas nuevas tecnologías y crear las sinergias para facilitar su participación en proyectos e iniciativas nacionales.

- Facilitar la organización sostenida de seminarios científico-tecnológicos para la actualización de nuestro entorno y una visión prospectiva sobre la proyección de futuro de estas tecnologías.

Al efecto, se pone en conocimiento la propia publicación asociada a la temática [12].

Referencias

- [1] Techopedia: Educating IT Professionals To Make Smarter <https://www.techopedia.com/>
- [2] Webopedia: Online Tech Dictionary for Students, Educators and IT Professionals, <https://www.webopedia.com/>
- [3] MIT Technology Review, <https://www.technologyreview.com/>
- [4] Ericsson, <https://www.ericsson.com/en>
- [5] Bodafone, <https://www.vmware.com/content/microsites/possible/stories-uk/vodafone.html>,
- [6] Blue Planet Technology, <https://www.blueplanet.com/technology/>,
- [7] China Mobile, <https://www.chinamobileltd.com/en/global/home.php>
- [8] Catalyst Digital Showcase, <https://inevent.com/en/TMForum-1588605874/catalyst-digital-showcase/hotsite.php>
- [9] China Telecom Corporation Limited, <https://www.chinatelecom-h.com/en/global/home.php>
- [10] Global Architecture Forum, <https://www.tmforum.org/global-architecture-forum/>
- [11] Orange, <https://www.orange-business.com/en/solutions/connectivity>
- [12] Pérez, Santiago; Facchini, Higinio; Dantiacq, Alejandro, Roberti, Bruno; Experimental Performance Contrast Between SDN and Traditional Networks, IEEE CHILECON, Santiago, Chile, 06 al 09 de diciembre de 2021.

Tecnologías de Smart IoT y aprendizaje automático para la solución de problemas en el medio productivo

Jorge Osio^{1,2}, Juan Salvatore¹, Mauro Salina¹, Marcelo Cappelletti^{1,2}, Diego Montezanti^{1,3}, Nicole Denon¹, Santiago Doti¹, Lucas Olivera¹, Chritian Botta¹, Matías Busum Fradera¹, Facundo Chazarreta¹, Daniel Terceros Quiroz¹, Diego Encinas^{1,3}, Martín Morales^{1,4}

¹ Programa TICAPPS, Univ. Nac. Arturo Jauretche, Florencio Varela (1888), Argentina.

² Grupo de Control Aplicado, Instituto LEICI (UNLP-CONICET), La Plata (1900), Argentina.

³ Instituto LIDI - Fac. de Informática -UNLP.

⁴ Centro UTN CODAPLI-FRLP, La Plata (1900), Argentina.

{josio, mcappelletti, dmontezanti, jsalvatore, mdsalina, dencinas, martin.morales}@unaj.edu.ar

Resumen

El presente proyecto se basa en la utilización de internet de las cosas (IoT) como herramienta fundamental para proveer soluciones a problemáticas de interés social, como lo es el cuidado del medioambiente y la innovación en el sector productivo, focalizando la investigación en las técnicas de aprendizaje automático, es decir, Smart IoT.

Entre los temas de investigación que se desarrollarán, se incluye el diseño e implementación de técnicas de visión por computadora con el objeto de agregar funcionalidades a dispositivos robóticos, de manera de proveer autonomía para determinadas tareas, con el agregado de control y supervisión remota mediante IoT. En esta línea, también se implementarán técnicas de visión por computadora para la clasificación de residuos reciclables mediante algoritmos de aprendizaje automático. Además, mediante técnicas de aprendizaje profundo y visión por computadora, se propone la clasificación de diferentes condiciones de cielo como consecuencia de la cobertura de nubes, lo cual será de suma utilidad para la optimización de sistemas que aprovechen la energía solar.

En esta propuesta se continúa con algunas líneas de procesamiento digital de imágenes con el agregado de técnicas de aprendizaje automático. Por otro lado, Teniendo en cuenta que las técnicas de procesamiento de imágenes aplicadas a visión por computadora requieren alto poder de cómputo, se considera necesario investigar la tolerancia a fallos del sistema de

procesamiento utilizado, para asegurar la correcta ejecución de dichos algoritmos.

En la misma línea de Smart IoT, se incluye en la propuesta actual el procesamiento y análisis de datos obtenidos de una red de sensores basada en IoT, lo que permitirá mediante técnicas de aprendizaje automático la implementación de un sistema de ayuda a la toma de decisiones, para optimizar y mejorar el aprovechamiento de los recursos agrícolas.

Palabras clave: *Aprendizaje automático, procesamiento de imágenes, Smart IoT, Visión por computadora, tolerancia a fallos en HPC.*

Contexto

Las líneas de I/D presentadas en este trabajo pertenecen al grupo SIoT&IA y están incluidas dentro del Programa TICAPPS (TIC en aplicaciones de interés social) de la Universidad Nacional Arturo Jauretche (UNAJ), Resolución N° 064/17, bajo la dirección del Dr. Ing. Martín Morales.

1. Introducción

En los últimos años ha surgido el concepto de Smart IoT (SIoT), favorecido por las mejoras de los procesadores para sistemas embebidos que los hacen aptos para el entrenamiento y ejecución de algoritmos de aprendizaje automático [1]. Esto permite obtener grandes beneficios cuando los dispositivos almacenan y analizan sus propios datos, de manera que

luego puedan comunicar esos datos a otros dispositivos conectados.

Para explotar las ventajas de las nuevas tecnologías, queda de manifiesto que internet de las cosas inteligentes es necesario para el almacenamiento y uso de grandes cantidades de datos en la nube, por otro lado, el aprendizaje automático permite el entrenamiento de máquinas para la realización de tareas y la visión por computadora permite la manipulación de objetos para automatizar actividades y procesos.

Dentro de la Visión por Computador, la detección de objetos es uno de los temas más candentes. Para entender el problema hay que pensar en cómo queda codificada una imagen digital. En general, para una máquina las imágenes son enormes cajas tridimensionales llenas de valores. Cada píxel (o punto) de la imagen queda representado con tres valores, que modifican su color. Así pues, cuando una máquina busca un objeto dentro de una imagen lo que realmente hace es buscar patrones que se correspondan con el objeto en particular.

Si dotamos a estas herramientas inteligentes con tecnologías como visión 3D, imágenes multiespectrales y lo combinamos con aprendizaje automático, para un análisis de datos más profundo y una predicción de modelos más precisa, conseguiremos una inteligencia artificial que dote de mayor flexibilidad a los equipos electrónicos para la toma de decisiones. De esta manera, se desarrolla la capacidad de configurar un robot a través del aprendizaje en lugar de usar programación prediseñada.

El concepto de internet de las cosas ha adquirido gran relevancia en los últimos años, debido a la posibilidad que ofrece de interconectar objetos entre sí y la conectividad a internet que provee a las redes de objetos [2]. Aunque los sistemas IoT son más que conocidos, en los últimos años ha surgido el concepto de Internet de Objetos Inteligentes (SIoT), que promete importantes avances en diferentes áreas de aplicación. Para este tipo de aplicaciones se dispone de un conjunto de protocolos de comunicación e interfaces que se pueden implementar mediante herramientas de sistemas embebidos como las desarrolladas en

[3-5]. Las investigaciones realizadas mediante SIoT se podrán aplicar en el automatismo y monitoreo de procesos, almacenamiento masivo de datos de sensores y dispositivos de cómputo (Big Data).

Respecto al aprendizaje automático [6], una de las áreas en donde se avanzó notablemente es en la de detección de objetos y clasificación de imágenes. Esto se debe en su mayor parte al desarrollo de nuevas técnicas de Machine Learning (Aprendizaje Automático) como el Deep Learning o Aprendizaje Profundo [7], además de las innovaciones en el manejo de Big Data (datos a gran escala) y el aumento en la capacidad de cómputo mediante el uso de diferentes tecnologías como cloud computing (computación en la nube) para el análisis de información. Algunos ejemplos de dichos avances en el área se aplican en: control de vehículos autónomos, detección de rostros, detección de matrículas, diagnóstico de enfermedades, realidad aumentada, etc.

Machine Learning es un subcampo de la IA en el que se utilizan diferentes algoritmos para recolectar datos, y con estos realizar un aprendizaje para luego hacer una predicción o sugerencia sobre algo [8]. De esta manera se permitirá resolver problemas de forma intuitiva y automatizada, sin que el mecanismo de elección se encuentre previamente programado. Por su parte, Deep Learning es un subcampo de Machine Learning, pero existen técnicas de Machine Learning que no utilizan Deep Learning. Este último es utilizado para realizar procesos de Machine Learning empleando redes neuronales artificiales compuestas por varios niveles jerárquicos [7]. En el nivel inicial la red aprende patrones simples, y esta información se envía al siguiente nivel de la jerarquía. Este segundo nivel toma la información obtenida en el primero y la combina con nuevos patrones aprendidos en este, generando información un poco más compleja, la cual es pasada al siguiente nivel, y así sucesivamente. Las técnicas de machine learning y deep learning proveen gran soporte para el diseño de aplicaciones de visión por computadora, es por eso que son parte fundamental de la propuesta.

Por último, teniendo en cuenta que algunos algoritmos serán ejecutados en sistemas de cómputo de altas prestaciones y la complejidad incorporada al implementar sistemas multi-cores, incrementa la vulnerabilidad a los fallos transitorios, estos fallos podrían corromper los resultados de las aplicaciones [9]. El alto costo (en términos temporales y de utilización de recursos) que implica volver a lanzar la ejecución de una aplicación desde el comienzo, en caso de que un fallo transitorio produzca la finalización de la aplicación con resultados incorrectos, justifica la necesidad de desarrollar estrategias específicas para mejorar la confiabilidad y robustez en sistemas de múltiples procesadores [10-12].

Aplicaciones

Entre las aplicaciones que se están desarrollando mediante IoT, procesamiento de datos, imágenes y machine learning se tiene; la detección automática de residuos reciclables, la detección de malezas, técnicas de eficiencia energética, calidad del aire, obtención de magnitudes climatológicas, radiación solar [13], visión artificial aplicada a robótica y la clasificación de diferentes condiciones de cielo.

El reciclado de desechos es considerado de suma importancia para el cuidado del medioambiente, debido a que, supone la reutilización de elementos u objetos ya utilizados, los que de otro modo serían desechados contribuyendo al incremento de la basura y al daño ambiental permanente. Gran parte del problema radica en el esfuerzo que requiere clasificar y separar los residuos inorgánicos, es por eso que en esta línea de investigación se propone desarrollar un sistema basado en redes neuronales artificiales que permita detectar y clasificar los objetos reciclables más comunes como papel, cartón, botellas, latas, etc, y los materiales con los que están hechos (plástico, vidrio, metal, papel) [14].

Respecto a la visión por computadora [15], se está llevando adelante la implementación de un sistema robótico que sea capaz de detectar obstáculos e identificar señales, mediante

procesamiento de imágenes, para dotarlo con capacidades autónomas. Además de un brazo robótico que permite seleccionar objetos previamente identificados mediante visión artificial [16].

Adicionalmente, se ha desarrollado un sistema de control para el sensado de gases y el control automático de iluminación y ventilación, de tal forma de contribuir con la eficiencia energética y calidad del aire y calidad del aire en ambientes cerrados [17].

Por otro lado, se implementó una red de sensores IoT que permiten obtener valores de más de 10 variables climáticas, como presión atmosférica, humedad del suelo, humedad ambiente, temperatura, intensidad lumínica, niveles de dióxido de carbono, entre otras. Lo que permitirá la implementación de un sistema inteligente que logre optimizar las plantaciones de cultivos en la zona.

Finalmente, la cobertura de nubes es un factor que atenúa y ocasiona intermitencia en los valores de la energía proveniente del sol. La aparición de grandes nubes que tapan el disco solar produce cambios repentinos en los valores de la radiación solar. Esta variabilidad provoca que la producción de energía que pueden generar los sistemas que aprovechen esta fuente como recurso energético no sea constante ni totalmente predecible en el tiempo. Sobre este punto, se ha comenzado recientemente el desarrollo de un sistema capaz de clasificar, mediante técnicas de aprendizaje profundo y visión por computadora, diferentes condiciones de cielo como consecuencia de la cobertura de nubes, lo cual será de suma utilidad para sistemas que utilicen la energía solar.

2. Líneas de Investigación y Desarrollo

El grupo de investigación que se ha constituido en la UNAJ es multidisciplinario, y sus miembros cuentan con experiencia en sistemas de cómputo de altas prestaciones, tolerancia a fallo, procesadores embebidos, IoT, aprendizaje automático, y robótica.

En cuanto a la robótica [16], esta línea está en pleno desarrollo y se buscan aplicaciones innovadoras como la detección de objetos para reciclado y la capacidad de detección de señales y obstáculos para el funcionamiento autónomo mediante procesamiento de imágenes [18].

Las líneas de I/D que se presentan en este trabajo están basadas en el estudio y desarrollo de herramientas alternativas para el procesamiento de imágenes a partir de algoritmos de procesamiento, aprendizaje automático y visión por computador.

Temas de Estudio e Investigación

Los datos recopilados de la robótica y de los sensores relacionados con el cultivo pueden suponer grandes cantidades de información dependiendo de los sensores de los que disponga y de la integración con otros sistemas. Esto permitirá conseguir una mayor precisión en los sistemas de Machine Learning y, cuantos más sean los datos útiles con los que alimentemos el modelo, mejor será su aprendizaje y, por ende, sus predicciones.

Los aportes originales no solo tienen que ver con el manejo de la información, sino también con las aplicaciones innovadoras que se le puede dar al aprendizaje automático en visión por computadora y robótica, donde el ejemplo más claro se muestra en la posibilidad de disponer de un sistema de reciclaje automático. También se puede innovar respecto a la visión por computador para conseguir la realización de tareas específicas mediante dispositivos robóticos en procesos particulares de la industria local.

Respecto a la tolerancia a fallos, partiendo del sistema desarrollado previamente, se buscará mejorar la eficiencia en la detección de fallos para que el tiempo demandado por esta tarea sea despreciable respecto al tiempo de procesamiento del algoritmo.

En cuanto al análisis de los datos climáticos en zonas de cultivos, nos permitirá estudiar cuáles de las técnicas de aprendizaje automático logra adecuarse mejor para determinar posibles problemas, y por ende soluciones en cultivos futuros.

Finalmente, se espera poder determinar la atenuación de la radiación solar global, a partir de la clasificación de diferentes condiciones de cielo (despejado, parcial o totalmente nublado) y de variables meteorológicas de sencilla adquisición.

3. Resultados y Objetivos

Resultados alcanzados:

La detección y clasificación de objetos reciclables son líneas de estudio comenzadas recientemente, aunque, ya se cuenta con la implementación de los primeros modelos basados en la arquitectura de redes neuronales convolucionales (CNN). Estos modelos se desarrollaron en un entorno virtualizado obteniendo como resultado tasas de acierto cercanas al 75%.

Por otro lado, se continuaron con las pruebas desarrollando modelos que hacen uso de las técnicas de aprendizaje por transferencia (Transfer Learning), en donde se logró una mejora en los modelos iniciales elevando la tasa de acierto a un 80% para los modelos entrenados desde cero y casi un 92% en modelos basados en Transfer Learning. Además, se espera a lo largo del año realizar pruebas en distintos microcontroladores para comparar resultados.

Respecto a la implementación del sistema Smart IoT, se desarrolló una aplicación en un miniordenador Raspberry Pi 3 Model B+, en donde se hace uso del módulo de la cámara (pi camera) para tomar fotos en tiempo real y realizar la clasificación de dicha imagen determinando que tipo de objeto reciclable se encuentra en ella [14]. Al respecto se está evaluando el porcentaje de acierto y los tiempos de predicción. La captura de la imagen y posterior clasificación arrojó tiempos aproximados a los 10 segundos. Estos se pretenden mejorar usando una Raspberry Pi 4 con 8 GB de RAM.

En cuanto a las aplicaciones en robótica, actualmente se está trabajando con un robot móvil con cámara y un brazo robótico. Respecto al robot móvil, se consiguió

implementar el guiado autónomo mediante detección de líneas en imágenes usando procesamiento en tiempo real, este método tiene problemas en las curvas pronunciadas, por lo que deberá combinarse la técnica con otro algoritmo de control.

Por otra parte, se logró implementar una red sensores basado en IoT que permiten obtener información de magnitudes climatológicas, del estado del suelo y otras con el fin de a partir de la información obtenida generar alertas tempranas que serán enviadas a los usuarios. Esta información permitirá perfeccionar las técnicas de cultivo y optimizar la cosecha. En esta línea de trabajo se espera a lo largo del año realizar mejoras en el sistema tanto a nivel de hardware como de software.

Por último, se debe destacar que, en cuanto a la tolerancia a fallos, se concluyó con los estudios de doctorado del Profesor Montezanti, obteniendo muy buenos resultados en sistemas de cómputo de altas prestaciones. Respecto a esto, se continúa con la línea en el nuevo proyecto de investigación.

Objetivos esperados:

- Promover la generación de conocimiento y el desarrollo de instrumentos que puedan aplicarse en la formación de recursos humanos para investigación.
- Fortalecer la actividad de investigación y vinculación que contribuya a remediar las problemáticas del territorio de influencia de la UNAJ.
- El objetivo principal de esta línea de investigación es desarrollar nuevas técnicas y obtener resultados favorables respecto a la detección de objetos y características del ambiente aplicando técnicas de visión por computadora y aprendizaje automático.
- Realizar el diseño y desarrollo de una metodología que permita tolerar fallos transitorios que se producen en las arquitecturas multicore (sistemas de múltiples procesadores), y que afectan especialmente la ejecución de aplicaciones paralelas de cómputo intensivo.

Se espera contribuir con el cuidado del medio ambiente, a través de la clasificación automática de objetos reciclables y la mejora de la producción agrícola mediante el análisis de variables climáticas a través de diferentes técnicas de Smart IoT.

4. Formación de Recursos Humanos

Uno de los principales objetivos del Programa TICAPPS, dentro de la temática de las líneas de I/D presentadas en este trabajo, es la formación de recursos humanos, tanto de docentes investigadores como de estudiantes.

Dentro de la temática de la línea de I+D, todos los miembros del proyecto participan en el dictado de asignaturas de la carrera de Ingeniería Informática de la UNAJ.

En este proyecto existe cooperación a nivel Nacional. Hay dos Doctores en Ingeniería, un Magister, un Especialista, cuatro Doctorandos y un integrante realizando su Maestría en temas relacionados.

Además, hay 3 estudiantes avanzados realizando las PPS de final de carrera en la temática. Actualmente, se encuentran en curso tres Tesis de Doctorado y tres becas de Estímulo a las Vocaciones Científicas del Consejo Interuniversitario Nacional (Becas EVC – CIN), relacionadas directamente con las líneas de I/D presentadas.

5. Bibliografía

[1]Mahmut Taha Yazici,Shadi Basurra and Mohamed Medhat Gaber, "Edge Machine Learning: Enabling Smart Internet of Things Applications". School of Computing and Digital Technology, Birmingham City University, Birmingham, UK

[2]Leila Fatmasari Rahman, "Choosing your IoT Programming Framework: Architectural Aspects", 2016 IEEE 4th International Conference on Future Internet of Things and Cloud

[3] J. Osio, M. Cappelletti , G. Suárez, L. Navarro, F. Ayala, J. Salvatore , D. Alonso ,D.

Encinas, M. Morales, “Diseño de aplicaciones de IoT para la solución de problemas en el medio socio productivo”, UNSJ, San Juan, WICC 2019.

[4] Jorge Rafael Osio, Juan Eduardo Salvatore, Mauro Salina, Diego Miguel Montezanti, Nicole Denon, Santiago Doti, Lucas Olivera, Matías Busum Fradera, Daniel Alonso, Marcelo Angel Cappelletti, Diego Encinas, Martín Morales. “Tecnologías de IoT y aprendizaje automático para la solución de problemas en el medio productivo y el cuidado del medioambiente”. XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021, Chilecito, La Rioja), 2021.

[5] J. Osio, J. Salvatore, D. Alonso, V. Guarepi, M. Cappelletti, M. Joselevich, M. Morales, “Tecnologías de la información y las comunicaciones mediante IoT para la solución de problemas en el medio socio productivo”, UNNE, Ciudad de Corrientes, WICC 2018.

[6] J. Hurwitz, D. Kirsch, “Machine Learning for Dummies”, Ed. Wiley, 2018.

[7] Nikhil Buduma, “Fundamentals of Deep Learning”, Editorial O’reilly. 2017

[8] Aurelien Gerón, “Hands-On Machine Learning with scikit Learn & TensorFlow”, Editorial O’reilly, 2017

[9] D. Montezanti, A. De Giusti, M. Naiouf, J. Villamayor, D. Rexachs, E. Luque, “AMethodology for Soft Errors Detection and Automatic Recovery”, in Proceedings of the 15th International Conference on High Performance Computing & Simulation (HPCS). ISBN: 978-1-5386-3250-5/17. IEEE, 2017, pp. 434

[10] J. Osio, J. Salvatore, E. Kunysz, D. Montezanti, D. Alonso, M. Morales, “Análisis de Eficiencia en Arquitecturas Multiprocesador para Aplicaciones de Transmisión y Procesamiento de Datos”, ITBA, CABA, WICC 2017

[11] J. Osio, D. Montezanti, E. Kunysz, Morales M., “Análisis de eficiencia y tolerancia a fallo en Arquitecturas Multiprocesador para aplicaciones de procesamiento de datos”, UNNE, Corrientes, WICC 2018.

[12] J. Osio, D. Montezanti, E. Kunysz, Morales M., “Determinación de la eficiencia y Estrategias de Tolerancia a Fallos en Arquitecturas Multiprocesador para aplicaciones de procesamiento de datos”, UNSJ, San Juan, WICC 2019.

[13] L. Olivera, J. Atia, L. Amet, J. Osio, M. Morales, M. Cappelletti, “Uso de redes neuronales artificiales para la estimación de la radiación solar horaria bajo diferentes condiciones de cielo. Avances en Energías Renovables y Medio Ambiente 24, 232-243, 2020/12/30

[14] Aprendizaje automático aplicado al procesamiento de imágenes para la clasificación de objetos reciclables M. Salina, J.R. Osio, M.A. Cappelletti, M. Morales. XXVII Congreso Argentino de Ciencias de la Computación (CACIC)

[15] Ashwin Pajankar, Raspberry Pi Computer Vision Programming Second Edition, (2015).

[16] Mohammad Jamshidi y Patrick J. Eicker, Robotics and remote system for hazardous environments, Prentice Hall, New Jersey, Estados Unidos (1993)

[17] Lo E., Fain F., Osio J., Cappelletti M., Aróztegui W., “Machine learning aplicada a variables ambientales”, VII CONAIISI, 2020.

[18] Wilhelm Burger & Mark J. Burge, “Digital Image Processing”, Second Edition (2008).

Estudio de la implementación del protocolo de Internet versión 6 en el ámbito regional y su impacto en los usuarios finales

Ernesto Sánchez, Álvaro Ignacio Gamarra, Daniel Arias Figueroa.

Universidad Católica de Salta / Facultad de Ingeniería /
Universidad Nacional de Salta / C.I.D.I.A. (Centro de Investigación y Desarrollo en
Informática Aplicada) / Facultad de Ciencias Exactas.
Campus Castañares, 08105558227 / Av. Bolivia 5150, 3874255408
esanchez@cidia.unsa.edu.ar, daaf@cidia.unsa.edu.ar, alvaroig@cidia.unsa.edu.ar

Resumen

Desde que, en el año 2020, el Registro de Direcciones de Internet de América Latina y Caribe (LACNIC) anunciara que ha otorgado la reserva del último bloque disponible de direcciones IPv4, hemos entrado en una cuenta regresiva hacia el proceso de migración a la “Era IPv6”. El proceso de transición será lento y por varios años ambas versiones del protocolo deberán coexistir. Si bien el trabajo técnico relacionado con el protocolo IPv6, en gran medida se ha completado, lo que resta mayoritariamente es su despliegue en las redes de los proveedores de servicios de Internet.

El presente trabajo expone los aspectos generales del Proyecto de Investigación “Estudio de la implementación del protocolo de Internet versión 6 en el ámbito regional y su impacto en los usuarios finales”, a desarrollarse en el ámbito de la Universidad Católica de Salta. A partir de los resultados a obtener, se espera, saber cuál es la situación en cuanto al despliegue del protocolo IPv6 en la región norte de nuestro país, conocer cuáles son los aspectos que demoran la implementación de dicho protocolo en los principales proveedores de servicios de

Internet y fundamentalmente, iniciar las acciones que acompañen el despliegue final del mismo.

Palabras clave: IPv6, ISP, Convergencia.

Contexto

La línea de investigación se encuentra apoyada por el Consejo de Investigaciones de la Universidad Católica de Salta. Para llevar adelante las tareas de investigación se cuenta con la infraestructura provista por el C.I.D.I.A. (Centro de Investigación y Desarrollo en Informática Aplicada) que depende de la Facultad de Ciencias Exactas de la Universidad Nacional de Salta. Se cuenta con el financiamiento del Consejo de Investigación de la Universidad Católica de Salta, por lo tanto, se disponen de todos los recursos necesarios para el desarrollo del proyecto.

Introducción

En agosto del 2020, el Registro de Direcciones de Internet de América Latina y Caribe (LACNIC) anuncia que ha otorgado la reserva del último bloque disponible de direcciones IPv4. Este agotamiento imposibilita, el crecimiento

de los usuarios de manera sostenible y principalmente la asignación de IPv4 públicas a usuarios finales. Cuando se habla de agotamiento IPv4, en LACNIC se refieren a una etapa de reservas en la que las asignaciones son restringidas en tamaño y periodicidad, la aplicación de estrategias de recuperación de direcciones no usadas, el reúso de las asignadas para otros propósitos y el uso de Network Address Translation (NAT) [1].

La creciente demanda por parte de usuarios finales, generada por la necesidad de conectar dispositivos tales como smartphones, smartTVs, entre otros, y principalmente tecnologías emergentes como IoT (Internet of Things), empujan a una solución de migración a IPV6 lo antes posible.

Por todo esto es que diversas organizaciones iniciaron acciones para implementar IPv6. Y no solo a usarlo como un complemento de IPv4, sino que llegue realmente a reemplazarlo por completo. Particularmente en nuestro país, se ha creado una Coalición IPv6, impulsada por Secretaría de Tecnologías de la Información y las Comunicaciones de la República Argentina [2].

El protocolo IPv6, surge como una evolución necesaria de la versión IPv4. Técnicamente, el espacio de direcciones de 32 bits de IPv4, permite hasta un máximo de 4.000 millones de direcciones, mientras que IPv6 tiene un espacio de direcciones de 128 bits, lo que equivale a 340 sextillones de direcciones. Además del mayor espacio de direcciones, IPv6 cuenta con otros beneficios;

- Seguridad: IPv6 incluye IPsec, que permite autenticación y encriptación del propio protocolo base.
- Multicast: Las especificaciones de IPv6 permiten la transmisión de paquetes de datos a múltiples destinatarios en una sola operación.
- Autoconfiguración: a través de SLAAC (“Stateless address autoconfiguration”) los hosts IPv6 pueden configurarse a sí mismos en forma automática, cuando se conectan a una red IPv6.
- Movilidad: A diferencia de IPv4 móvil, IPv6 móvil evita el ruteo triangular, y por ello es tan eficiente como el IPv6 nativo [3].

Son varios los factores que impiden un rápido despliegue de IPv6, el principal factor está relacionado con la compatibilidad, ya que IPv6 no fue diseñado para que sea compatible con IPv4, lo que significa que las redes IPv6 no pueden comunicarse de manera transparente con redes IPv4. Dada la incompatibilidad entre ambos protocolos, estos deberán coexistir durante un período más o menos prolongado, provocando que el proceso de transición sea gradual. Colateralmente, se deben resolver otras implicancias de nivel técnico relacionadas con otros protocolos de Internet, el Sistema de Nombres de Dominio, Calidad de Servicio, Seguridad, y otros aspectos derivados por la implementación de un entorno de Dual Stack, (IPv4 e IPv6) [4]. En base a lo expresado anteriormente, se hace necesario el uso de técnicas de transición que permitan mantener la conectividad tanto de IPv4, como de IPv6. Estas técnicas deben aplicarse tanto del lado de los dispositivos terminales, los cuales deberán decidir cual Stack de protocolo usar al momento de solicitar servicios de conexión de red, por otro lado, los dispositivos del núcleo de la red, como switches, routers, firewalls, deberán

expandir su hardware y software para soportar las características de IPv6.

Líneas de Investigación, y Desarrollo

Los principales ejes temáticos que se están investigando son los siguientes:

- Protocolo de Internet IPv6.
- Convergencia IPv4 IPv6.
- Internet.
- Simulación.

Resultados Obtenidos/Esperados

Para llevar adelante las actividades inherentes al proyecto de investigación se definieron tres etapas principales:

Recopilación Documental: Con el propósito de extraer información de diferentes documentos, libros, papers, etc., se utilizará la metodología de Estudio de Mapeo Sistematizado (Systematic Mapping Study) [5], muy empleado en diferentes áreas de investigación, lo cual permitirá evaluar cual es el grado de avance de la implementación y despliegue del protocolo IPv6 en nuestro país, focalizando en la región norte de nuestro país, donde se realizarán entrevistas a los principales proveedores de Internet a fin de conocer aspectos que demoren el despliegue y adopción de IPv6.

Diseño e implementación de escenarios de pruebas basados en herramientas de simulación que permitan evaluar el desempeño de protocolos y servicios de Internet, tales como el Sistema de Nombres de Dominio, Calidad de Servicio, Seguridad, Enrutamiento, etc., en un entorno de Dual Stack, (IPv4 e IPv6). Cabe destacar que, en el marco de proyectos anteriores, se realizaron pruebas

de despliegue e implementación de IPv6 sobre el protocolo de enrutamiento BGP, alcanzado los objetivos propuestos [6].

En la etapa final, y con los resultados obtenidos de las etapas anteriores, se proponen actividades de transferencia al medio, convenios otras instituciones educativas a fin de conformar grupos de trabajo para el despliegue final del protocolo IPV6.

Formación de Recursos Humanos

El grupo de investigación conformado se caracteriza por una constitución heterogénea de profesionales vinculados a la informática. El director es Doctor en Ciencias Informáticas por la UNLP y el Codirector Master en Redes de Datos. Dentro de los investigadores se cuenta con una Magister en Ingeniería de Software, un Ingeniero en Sistemas y un Técnico Universitario en Programación. También integran el grupo alumnos avanzados de la carrera Ingeniería en Telecomunicaciones de la Universidad Católica de Salta. En el transcurso del proyecto se tiene como objetivo consolidar la formación en investigación de los integrantes de menos antecedentes y también está contemplado que uno de los integrantes complete el cursado de la Maestría en Redes de Datos y la certificación Mikrotik.

Bibliografía

- [1] Investigación sobre despliegue de IPv6 en América Latina y Caribe 2020. LACNIC 35. <https://www.lacnic.net/innovaportal/file/5>

229/1/smc+lacnic---ipv6-lacnic-35-
20210513-v1-03.pdf

[2] IPv6 Benchmark Regional Coalición IPv6.

<https://www.argentina.gob.ar/jefatura/innovacion-publica/ssetic/grupo-de-trabajo/ipv6>

[3] PROYECTO PMIP6: Análisis, Evaluación y Comparación de ambientes Proxy Mobile IP en versión 6, aplicado a Redes de Avanzada. Departamento de Electrónica - UTN Facultad Regional Mendoza.

http://sedici.unlp.edu.ar/bitstream/handle/10915/19966/Documento_completo.pdf?sequence=1

[4] Técnicas para el despliegue de IPv6 en redes LAN. Universidad Católica de Santiago del Estero.

<http://sedici.unlp.edu.ar/handle/10915/76964>

[5] B. A. Kitchenham. “Using mapping studies as the basis for further research - a participant-observer case study”

[6] Simulación de enrutamiento BGP con GNS3.WICC 2021.

<https://drive.google.com/file/d/154tPVFoKFGGe0WGATAPLAv0vc81RaPInt/view>.

Pág. 38 a 42.

Entorno de Contenedores de Sistemas Embebidos con Conexión a Dispositivos Externos

Waldo Valiente, Esteban Carnuccio, Mariano Volker, Matías Adagio, Micaela Antelo
Departamento de Ingeniería e Investigaciones Tecnológicas
Universidad Nacional de La Matanza
Dirección: Florencio Varela 1703 – CP 1754 –
{wvaliente, ecarnuccio, mvolker, matadagio, mantelo}@unlam.edu.ar

RESUMEN

Los Sistemas Ciber Físicos toman relevancia, con el surgimiento de tecnologías populares como Internet de las cosas, Industria 4.0, Internet industrial, máquina a máquina, Internet de todo y la capa Fog. Su contexto general de trabajo se basa en las interacciones dadas entre las funciones de control y los mecanismos de comunicación entre los componentes que forman el sistema. Esta interacción afecta al ambiente externo, que luego el sistema se adapta a ese cambio que él mismo produce. En la actualidad hay diferentes iniciativas y proyectos educativos que buscan enseñar estas tecnologías. Uno de los métodos de enseñanza es a través de prácticas con el sistema físico. Esto tiene problemas inherentes a la compra de los componentes y/o por la inexperiencia en la utilización de su electrónica. Como alternativa se busca que el uso de emuladores de sistemas embebidos permita enfocar directamente sobre el aprendizaje. Así facilita la tarea de abordar diferentes escenarios sobre los mecanismos de comunicación utilizando los protocolos Bluetooth y WiFi. Para sostener la infraestructura se utilizarán contenedores Docker, que ayudan en la construcción, gestión y pruebas de complejas topologías de sistemas.

Palabras clave: Sistemas Embebidos, Emulación, Comunicación.

CONTEXTO

Nuestra Línea de Investigación es parte del proyecto “*Entorno de Contenedores para sistemas embebidos con conexión a dispositivos externos*”, dependiente de la Unidad Académica del *Departamento de Ingeniería e Investigaciones Tecnológicas*, perteneciente al programa de Investigaciones CyTMA2 de la Universidad Nacional de La Matanza, el cual es formado por docentes e investigadores de la carrera de ingeniería en informática. Este proyecto es continuación de los trabajos que viene realizando el grupo de investigación en sistemas operativos y en el área de Internet de las cosas.

1. INTRODUCCIÓN

La noción de los Sistemas Ciber Físicos, denominación que proviene del inglés Cyber-Physical Systems (CPS). Surgió en el año 2006 por Helen Gill en su discurso en la Fundación Nacional de Ciencias de EE. UU. Aunque en [1] interpreta que esta definición posee mayor antigüedad. Ya que, establece que es un término que deriva de la definición de “cibernética”.

Cuyo origen proviene desde mediados del siglo XX, nacido en el estudio de sistemas complejos de control de retroalimentación. El contexto de trabajo de los CPS se centra en hallar las interacciones dadas entre las funciones de control y los mecanismos de comunicación entre los componentes del sistema. Esta interacción afecta al ambiente externo, luego el sistema se debe adaptar al cambio que él mismo produce. En la actualidad el término CPS toma nueva relevancia, con el surgimiento de tecnologías populares como Internet de las cosas (IoT), Industria 4.0, Internet industrial, máquina a máquina (M2M), Internet de todo (IT) y la capa Fog (una capa intermedia de procesamiento en la topología de computación en la nube). Estos reflejan una visión de una tecnología que conecta profundamente el mundo físico con la información necesaria para interactuar en el día a día. El término CPS es una herramienta fundamental y duradera, ya que permite unir a la ingeniería tradicional del mundo cibernético y físico con las personas sin importar la distancia.

En los últimos años creció de forma exponencial la tecnología de Internet de las Cosas. El término IoT toma relevancia cuando se superó la cantidad de dispositivos conectados a internet, que el número de personas que existían en el mundo en ese momento. Según las proyecciones recompiladas por [2], se estima que actualmente hay 75.000 millones de dispositivos conectados y en el año 2025 habrá aproximadamente 100 mil millones. Además, que el instituto McKinsey Global sugiere que el impacto financiero de IoT en la economía global puede ser de entre 3,9 a 11,1 billones de dólares para 2025. El potencial anticipado de IoT se puede realizar cuando las “cosas” de Internet comienzan a interactuar de forma automática e inteligente con otras “cosas” de Internet.

Algunas de esas “cosas” como las computadoras y los teléfonos inteligentes pueden conectarse directamente a Internet a través de Ethernet o protocolos de comunicaciones móviles estándar, mientras que otros dispositivos se conectarán a Internet a través de una compuerta de enlace o enrutador local. De hecho, la compuerta incluso podría ser un teléfono inteligente que se comunica con otros sensores y dispositivos en proximidad. Por lo que el SE es claramente una plataforma flexible, capaz de desempeñarse como “Cosa” y/o la compuerta de enlace, en la medida en que sea capaz de conectarse directamente a Internet a través de Ethernet o, para dispositivos portátiles y más pequeños, capaz de conectarse con la compuerta a través de Bluetooth o comunicación inalámbrica como WiFi [3].

Por otro lado, el objetivo primario de la virtualización es recrear un componente físico o hardware, mediante un programa (software), emulando así todas las funcionalidades de este. Se pueden virtualizar: dispositivos, redes, componentes hardware, sistemas operativos, servidores, entre otros. Las tecnologías de virtualización como QEMU permiten a los desarrolladores emular un dispositivo físico, así se puede comenzar el desarrollo antes de fabricar el hardware [4]. El emulador y visualizador de la máquina QEMU permite a los desarrolladores probar el dispositivo de forma segura, incluso permite corregir defectos que pueden bloquear todo el sistema. El desarrollo y depuración de programas o controladores dentro de un emulador, hace que sea una tarea similar al desarrollo de aplicaciones comunes. Ya que, en el peor de los casos, los errores pueden provocar que el emulador se bloquee, se corrige el error y se vuelve a iniciar la simulación, sin mayores contratiempos [5].

Hay varias iniciativas y proyectos educativos que buscan enseñar las tecnologías mencionadas, orientándolas a estudiantes universitarios y de pregrado. En [6] se recapitulan las diferentes formas de enseñanza de estos temas, diferencia a los enfoques principales para los tópicos de gestión de proyecto, el diseño del sistema o las técnicas de redes. Enfatiza que el uso de una plataforma flexible en donde se pueda ejecutar los Sistemas Embebidos (SE) con conectividad de red. La conexión es una característica vital, ya que los SE utilizan distintos esquemas que les permiten comunicarse con otros dispositivos externos. Para esto se utilizan mecanismos tales como WiFi, Bluetooth, Ethernet, entre otros. De esta manera pueden intercambiar datos con servidores u otros dispositivos. Para así formar una topología de computación en la nube [7] y [8]. Para gestar esta idea, los contenedores Docker son una de las herramientas que ayudan en la construcción, gestión y pruebas de complejas topologías de sistemas. Ellos brindan un ambiente aislado que permiten trabajar con paquetes de software sin utilizar virtualización del hardware. La principal diferencia del uso de contenedores con una máquina virtual. Es que está realiza una abstracción completa del dispositivo físico y necesita ejecutarse sobre un Sistema Operativo (SO) que es redundante. Mientras que los contenedores se ejecutan sobre el motor de contenedores, que a su vez ejecuta integrado al SO anfitrión, que lo emplea como base para poder funcionar y acceder a los recursos del equipo [9]. Gracias a sus virtudes tiene un impacto que va más allá de su tradicional origen, dedicado al desarrollo o pruebas de programas. Los contenedores de Docker se están extendiendo a ámbitos como la investigación científica o la comunidad docente [10] y [11].

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

La enseñanza de los sistemas embebidos tiene como dificultad que no existen herramientas de simulaciones con características completas de licenciamiento abierto. Esta dificultad surge porque las herramientas de acceso libre tienen limitados sus recursos. Las herramientas profesionales, se enfocan a un mercado donde prevalece el licenciamiento del tipo comercial. Sin embargo, ofrecen versiones destinadas a alumnos, pero muy limitadas. Estas limitaciones, tanto en las herramientas de acceso libre como las comerciales, entorpecen la tarea de la enseñanza plena. Por este motivo, en esta investigación, se plantea la construcción de un entorno de emulación, que permita la conexión del SE con sistemas externos. Formando un ecosistema integro mucho más amplio. Así se presentan pruebas sobre escenarios complejos de funcionamiento. De esta forma, se dispone de entornos listos para enseñar en forma completa sobre el amplio espectro del universo CPS. Esto se sustenta, en el hecho de que existen diferentes métodos de enseñanza. El más utilizado es el tradicional. El cuál consiste en realizar las prácticas con el sistema físico. Esto tiene problemas inherentes a la compra de los componentes y/o por la inexperiencia en la utilización de su electrónica. Como alternativa surge el uso de SE emulados. Esto permite enfocar las prácticas directamente sobre el aprendizaje, dejando de lado las problemáticas antes mencionadas. En este sentido existen diferentes simuladores, que presentan determinadas características y limitaciones, que dificultan el aprendizaje completo, ya que no están preparados para ello.

El simulador Thinkercad, es utilizado en la educación a distancia de sistemas embebidos

[12]. Esta es una plataforma web que permite simular el SE Arduino Uno. De tal manera que ofrece la posibilidad de conectarla con distintos sensores y actuadores virtuales [13]. Pero su falencia es que no presenta conectividad con el exterior. Ya que no permite emular un módulo de Bluetooth ni de WiFi. Por otro lado, se encuentra el simulador Proteus, que permite simular los componentes de Bluetooth y WiFi. Pero, no posee conexión con el exterior. Además, presenta una licencia del tipo comercial. Aunque brinda una versión para estudiantes, resulta bastante restringido su uso, por ejemplo, no permite guardar lo realizado, entre otras limitaciones [14]. Por otra parte, está la página Wokwi que permite simular algunas placas de desarrollo básicas [15]. Pero, la emulación de la conectividad del WiFi y Bluetooth no se encuentra implementada. Siguiendo la misma línea, existe un simulador específico para el SE Raspberry Pi, que ofrece Microsoft, desde su plataforma Azure [16]. Sin embargo, este proyecto se encuentra en una versión preliminar, ya que posee un único ejemplo con funcionalidades muy restringidas. Esto se debe a que no permite al usuario reemplazar sus componentes de hardware. Otro de los emuladores que existen es el Android Emulator, que está integrado en la herramienta Android Studio, que permite emular un dispositivo móvil incluyendo al SO Android [17]. La herramienta facilita las pruebas, dentro de un simulador, de la aplicación Android mientras se está desarrollando.

Por ende, se pretende desarrollar una herramienta de enseñanza y trabajo, que permita a la comunidad llevar a cabo sus proyectos de IoT en forma simulada. De forma tal, que permita construir e implementar sistemas simulados comunicándolos con dispositivos externos reales. Para ello se proyecta que se

podrá llevar a cabo, a través de algunas de las formas de comunicación antes mencionadas. Esto se realizará empleando emuladores, tales como Qemu, configurado dentro de contenedores. A fin de que pueda ser de rápido uso e instalación.

3. RESULTADOS OBTENIDOS/ESPERADOS

Se espera mejorar el proceso de aprendizaje y de pruebas sobre los CPS. A través de un único entorno modular. Que este permitirá la integración de distintos dispositivos formando topologías complejas de múltiples dispositivos. Permitiendo así emular el comportamiento de la plataforma completa. Todo el proyecto se espera que sea disponible y documentado sobre repositorios públicos, bajo la metodología de acceso abierto, permitiendo el uso a toda la comunidad científica y de alumnos.

4. FORMACIÓN DE RECURSOS HUMANOS

La presente línea de investigación, dentro del departamento de Ingeniería e Investigaciones Tecnológicas, forma parte del trabajo que uno de los investigadores se encuentra realizando para su maestría. Completan el grupo de investigación dos de docentes de categoría V y un ingeniero en formación de investigador.

5. BIBLIOGRAFÍA

- [1] E. A. Lee y S. A. Seshia, *Introduction to Embedded Systems: A Cyber-Physical Systems Approach*, Cambridge, Massachusetts, Estados Unidos: The MIT Press, 2017.
- [2] T. Al-Rousan, «The Future of the Internet of Things,» de *Journal of Computing*,

- Communications & Instrumentation Engg*, 2017.
- [3] R. Toulson y T. Wilmshurst, *Fast and Effective Embedded Systems Design: Applying the ARM Mbed*, 2 ed., Newton, MA: Newnes, 2012.
- [4] qemu, 09 2019. [En línea]. Available: <https://www.qemu.org/>.
- [5] A. Kotovsky, *How to Develop Embedded Software Using the QEMU Machine Emulator*, Wilmington, Estados Unidos: Apriorit Inc., 2019.
- [6] S. Martin, *Teaching and Learning Advances on Sensors for IoT*, Basel, Suiza: MDPI, 2021.
- [7] P. Waher, *Learning Internet of Things*, Packt Publishing, 2015.
- [8] Sachan, *Internet de las cosas (IoT) y sus aplicaciones*, 2020.
- [9] J. M. Ortega, *DOCKER. Seguridad y monitorización en contenedores e imágenes*, RCLibros, 2019.
- [10] C. Boettiger, «An introduction to Docker for reproducible research, with examples from the R environment,» *ACM SIGOPS Operating Systems Review, Special Issue on Repeatability and Sharing of Experimental Artifacts*, pp. 1-24, 2015.
- [11] D. N. Nüst, V. Sochat, B. Marwick, S. J. Eglen, T. Head, T. Hirst y B. Evans, «Ten Simple Rules for Writing Dockerfiles for Reproducible Data Science,» *PLOS Computational Biology*, pp. 1-24, 2020.
- [12] J. Lopez Cisneros y G. Torales, «Experiencia de transición de una clase presencial a virtual en tiempos de Covid-19: una visión desde el docente,» 2020.
- [13] Autodesk, «<https://www.tinkercad.com/>,» 2022. [En línea].
- [14] Labcenter, 2022. [En línea]. Available: <https://www.labcenter.com/downloads/>.
- [15] CodeMagic, 2019. [En línea]. Available: <https://wokwi.com/>.
- [16] Microsoft, 2022. [En línea]. Available: <https://docs.microsoft.com/es-es/azure/iot-hub/iot-hub-raspberry-pi-web-simulator-get-started>.
- [17] G. Developers, 2022. [En línea]. Available: <https://developer.android.com/studio/run/emulator?hl=es-419>.

Desarrollo de Dispositivos Electrónicos de Aplicación Específica a Baja Escala

Ing. Alejandro Fourcade, Mg Jorge Eterovic

Departamento de Ingeniería e Investigaciones Tecnológicas Universidad Nacional de La Matanza

Florencio Varela 1903 (B1754JEC), San Justo, (5411) 4480-8900

afourcade@unlam.edu.ar; eterovic@unlam.edu.ar

RESUMEN

Los métodos y procesos de desarrollo e implementación de circuitos electrónicos han sufrido cambios importantes en las últimas décadas. Desde la ubicación de los grandes centros de producción y de diseño, hasta las etapas que deben seguirse para completar exitosamente un proyecto electrónico, han sufrido cambios, correcciones y modificaciones sucesivas.

Analizando las plataformas de desarrollo embebido que ofrece el mercado, su creciente complejidad y especificidad, su disponibilidad a bajo costo y su gran cantidad de casos de uso documentados, se puede concluir que el escenario tecnológico para el programador de sistemas embebidos ha cambiado lo suficiente como para fijar nuevas reglas para desarrollos rápidos, técnicamente avanzados y altamente competitivos.

La Industria 4.0, los diferentes IOTs (Internet of Things): Industrial IOT, Enterprise IOT, automóviles autónomos, edificios inteligentes, robots industriales interconectados, son solo algunas de las aplicaciones tecnológicas que refuerzan el análisis de qué nivel de conocimiento específico debe tener un profesional hoy.

En cuanto a los objetivos conceptuales del trabajo de investigación, los cambios anteriormente descriptos dejan en evidencia la necesidad de actualizar la visión del acceso y aplicación del conocimiento. La transmisión de conocimientos enmarcados en un contexto de aplicación con un nivel de abstracción superior al nivel de bit.

El proyecto propone construir hardware digital, que incluya tecnología avanzada en microcontroladores y en DSP (Digital Signal Processing) de aplicación real en el mercado. Para

llevar a cabo este objetivo se utilizarán estrategias de dirección de proyectos que incorporen conceptos de métodos ágiles, MVP (Minimum Viable Product), ciclos de desarrollo iterado, mejora continua y la metodología COTS (Commercial off-the-self).

Palabras Clave: *Diseño electrónico; DSP – Digital Signal Processing; Software COTS – Commercial off-the-self.*

CONTEXTO

Este proyecto de investigación se desarrolla en el marco de un Programa de Incentivos a Docentes Investigadores de la Secretaría de Políticas Universitarias (PROINCE) en el Departamento de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de La Matanza.

El proyecto es financiado por el propio Departamento y es del tipo investigación aplicada. El mismo propone construir hardware digital que incluya tecnología avanzada en micro controladores y en DSP (Digital Signal Processing) de aplicación real en el mercado. Los trabajos de campo y relevamientos realizados aportaron información valiosa y sirvieron como base para el presente trabajo.

1. INTRODUCCIÓN

La drástica alteración de los ciclos de vida de los productos tecnológicos ha modificado las necesidades y objetivos de los procesos de desarrollo. El tiempo de puesta en mercado de las nuevas tecnologías debe ser lo más corto posible, en caso contrario se corre el riesgo de generar un producto con especificaciones desactualizadas.

La importancia del desarrollo intensivo, orientado a resultados y con objetivos operativos colabora con la disminución del TTM (Time To Market), parámetro crítico dentro de cualquier esquema de desarrollo tecnológico actual.

Por ello, los métodos clásicos de diseño e implementación han cambiado, acortando los plazos, aun los que corresponden a las pruebas de confiabilidad. La documentación, tanto de los procesos como de los productos se ha minimizado brindando información técnica más genérica y menos abundante.

La cultura que rodea a la generación de un producto ya sea diseño electrónico como programación de firmware y aplicaciones requiere hoy de reflejos rápidos para no comprometer las posibilidades de éxito [1]. El desarrollo de software de aplicación ha cambiado de enfoque en sus procesos y hoy conviven modelos con mandatos casi opuestos.

Mientras que en el desarrollo en cascada el proyecto se hace según la especificación inicial, en los métodos ágiles se puede corregir, ampliar o recortar los requerimientos de inicio [2]. En Scrum se realizan ceremonias que reafirman los puntos del manifiesto Ágil, se minimiza la documentación, se centra el desarrollo en la satisfacción del cliente y se itera en Sprints generando productos funcionales intermedios (MVP) [3].

En los métodos clásicos se prioriza llegar con el producto terminado a la fecha límite fijada al principio del proyecto, y evitar a toda costa excederse en plazos para no tener penalidades ni comprometer del presupuesto previsto.

Mientras que el software de aplicaciones evolucionó e incorporó cambios de proceso y principalmente, aceptó pensar diferente el proceso de desarrollo, en los sistemas embebidos o en los desarrollos electrónicos nada de eso sucedió, fundamentalmente porque el axioma, en un principio, fue que no podrían aplicarse métodos ágiles en desarrollos con software que maneja hardware [4].

Los procesos que intervienen en el diseño de productos embebidos siguen siendo hoy, mayoritariamente los tradicionales, pero lentamente muestran creciente permeabilidad a incorporar nuevos enfoques y procesos.

Las opciones disponibles a la hora de comenzar el desarrollo de un sistema embebido se han

diversificado, proponiendo nuevas variantes. Algunas son la evolución de herramientas preexistentes, otras formas más ágiles y sencillas de generar código. Mostrar resultados en la nube o en la web se ha transformado hoy en un objetivo fácilmente accesible gracias a plataformas, middlewares o protocolos específicos. Esa es una de las razones por las que se ha producido una revolución principalmente en las potencialidades de los desarrollos de baja y media escala [5].

Otra de las razones que impulsó este cambio fue la drástica disminución de los precios de los microcontroladores y su multiplicación de plataformas de desarrollo. Este hito, puso al alcance de muchos pequeños desarrolladores la potencia de microcontroladores avanzados. La posibilidad de prototipar a bajo costo, el acceso a impresoras 3D para diseñar gabinetes o accesorios, la oferta creciente de cortes laser o trabajos en CNC ponen al alcance del desarrollador de hardware embebido de baja escala posibilidades en otro momento impensadas.

Los componentes electrónicos discretos, que ayer eran la unidad funcional mínima a considerar en un diseño, se han reemplazado hoy por módulos funcionales y placas de desarrollo, convirtiendo los procesos de diseño electrónico en procesos de integración de partes [6].

La mención a la escala es relevante, ya que se ha facilitado el acceso a funciones avanzadas de microcontroladores o módulos de DSP desde plataformas de bajo costo y amplia difusión.

La posibilidad de implementar sistemas operativos en tiempo real (RTOS), depuración sobre el producto terminado, disponibilidad de código reusable en conjunto con la disminución de los precios del hardware ha puesto a disposición de la pequeña industria el acceso a los grandes ambientes de desarrollo [7].

La integración de hardware y software COTS (Commercial off-the-shelf) o subconjuntos armados de amplia distribución y disponibles al público en general, ha resultado sumamente exitosa en dispositivos comerciales [8]. Este concepto que proviene de la defensa militar y ayudó a bajar costos operativos y optimizar tiempos de aprovisionamiento de insumos no críticos.

La integración de productos que puedan adquirirse masivamente para armar diseños y

luego consolidarlos en un prototipo comienza a popularizarse.

En el ámbito comercial, se encuentran implementaciones exitosas realizadas con estos preceptos. Desde la puesta en órbita del satélite Fossasat-1, hecho en gran parte de hardware COTS utilizando módulos LORA de bajo costo, a la integración exitosa de las impresoras 3D, tomando como caso de éxito a la empresa Prusa con más de 100.000 unidades vendidas que en sus comienzos generó modelos altamente exitosos con hardware genérico y firmware abierto.

El desarrollo de investigación en el ámbito de la integración como herramienta y la satisfacción al cliente como norte, exige la generación de experiencias de aplicación, recolección de ideas e impresiones para acordar las mejores vías de planificación, implementación y evaluación de resultados.

2. LÍNEAS DE INVESTIGACION Y DESARROLLO

En el escenario tecnológico actual existen tendencias que muestran la necesidad de modernizar los marcos conceptuales. Desde el punto de vista académico, la enseñanza imparte las nociones técnicas que en muchos casos se ven disociadas de su aplicación práctica. Eso hace que el alumno pierda interés en incorporar conocimiento ya que no detecta que le proporcione ventajas inmediatas.

Es por eso que es necesario asociar a los conocimientos técnicos su ámbito de aplicación, expresando a su vez las razones que impulsan a recorrer caminos de superación y cambio. El marco conceptual del conocimiento, los motivos de su desarrollo, su evolución, los motores del cambio; forman parte del mismo conocimiento tecnológico ya que agregan valor y visión.

Los procesos tradicionales de desarrollo de sistemas embebidos se han mantenido sin mayores cambios desde hace décadas. Los lenguajes de programación utilizados siguen siendo, en su mayoría, los mismos desde hace tiempo (C, C++, Assembler). Aunque algunas otras opciones comiencen a ganar terreno (Java, Python), principalmente hoy se utiliza C como herramienta de programación resultando la elección más

frecuente por su alto rendimiento y amplia variedad de bibliotecas.

Las circunstancias de mercado y la expansión en la base de programadores de embebidos han forzado la reedición de una antigua comparación entre lenguajes compilados e interpretados. La performance dejó de ser el argumento principal y pasó a ser un argumento más, junto con la portabilidad, la curva de aprendizaje, la interfaz visual y la disponibilidad de bibliotecas y soporte.

Como se mencionó anteriormente la imposibilidad de innovación en cuanto a los caminos de desarrollo se ve limitada por la inclusión del hardware en el proceso de diseño. Históricamente en sistemas embebidos se desarrolló el hardware y luego el software, obligando a secuenciar los pasos y a depender de los tiempos que imponen la generación de un prototipo funcional de un sistema electrónico. Hasta hace pocos años era impensable modificar ese orden y las planificaciones lo incorporaban a sus fases de trabajo la del diseño e implementación de hardware.

Eso implicaba generar una secuencia que obligaba a seguir una sucesión de tareas como, por ejemplo: toma de requisitos, generación de especificaciones, generación de hardware, programación de software, generación de un prototipo funcional, pruebas, entrega al cliente y fase de mantenimiento.

Actualmente, con la incorporación de simuladores avanzados, la posibilidad de producir prototipos rápidamente (aun a baja escala), la aceleración de los tiempos de logística y la diversificación de proveedores son factores que impulsan a repensar los procesos, ya que está en juego el liderazgo tecnológico y la competitividad.

Existen grupos de desarrollo, conferencias anuales y numerosos trabajos de investigación que proponen la incorporación de conceptos ágiles al desarrollo de sistemas embebidos. Gran parte de las frustraciones y fracasos en esta línea de acción se debe a la voluntad de adaptar completamente los desarrollos embebidos a los métodos ágiles.

Las propuestas más conciliadoras, como por ejemplo el trabajo “Agile Development Methodology for Embedded Systems: A Platform-Based Design Approach” [9], eligen tomar y aplicar algunos principios de dos métodos ágiles (XP y Scrum) y advierte además que el proceso de

desarrollo de sistemas embebidos no es totalmente compatible con estos principios.

A pesar de esta aparente incompatibilidad, se puede articular y aplicar conceptos como reusabilidad de código, ciclos de programación iterativos, mejora de procesos y cambio de especificaciones durante el proyecto.

Desde lo académico, balancear la profundidad de los conocimientos con la diversidad que los comprende se hace indispensable. El crecimiento exponencial de los tipos de tecnologías, su complejidad y sus diversas aplicaciones obliga a replantear las bases y los objetivos del proceso de transmisión de conocimientos.

En el caso de los sistemas embebidos como unidad de conocimiento, es necesario hoy comprender sus funciones, potencialidades, aplicaciones, integraciones, protocolos de interconexión, interacción con sensores y actuadores, formas de comunicación inalámbrica, posibilidades de bajo consumo, en lugar de conocer a fondo los mapas de bits que controlan sus procesos internos.

Este salto en cuanto a unidad mínima de información se debe a la evolución y ampliación de los conocimientos que hoy son indispensables para tener una visión más amplia y completa de las etapas y partes de un proyecto tecnológico. Es necesario, sin perder concepto, subir un nivel en la capa de abstracción y priorizar el algoritmo por sobre los detalles de implementación.

Es así, que se hace mandatorio un replanteo tanto de las formas y las vías para el desarrollo de sistemas embebidos, como analizar, por ejemplo, cuáles son los conocimientos mínimos que son necesarios para tener una visión funcional de los proyectos tecnológicos que vaya más allá de lo netamente técnico.

Se utilizarán las instalaciones de la Universidad, incorporando el equipamiento adicional necesario. Se realizarán luego análisis y mediciones sobre equipos reales que oficiarán de patrones que permitan controlar la calidad del proceso y sus resultados.

El proceso de análisis desembocará en la determinación del equipamiento a producir, se comenzará una etapa de diseño mecánico que incluirá chasis, PCBs y componentes.

Una vez finalizada, se realizarán las simulaciones necesarias en software y las pruebas de rendimiento de prototipos en laboratorio.

Luego se definirán los pasos a seguir para producción del equipo a baja escala. Los fondos asignados al proyecto se destinarán a la adquisición de recursos y componentes.

El proceso entonces abarcará desde la etapa de análisis de requisitos hasta la realización física. El desarrollo se llevará a cabo con una metodología de desarrollo utilizada actualmente en situaciones de este tipo, se trata de la integración y generación de modelos mínimos que luego completarán sus funciones en ciclos de desarrollos asimilables al método Agile de producción de software.

3. RESULTADOS OBTENIDOS/ESPERADOS

Nuevos ángulos de visión necesitan, para materializar sus resultados, la generación de procedimientos y reglas que resultarán luego en la modificación de los procesos existentes. Lo que se desea investigar es la factibilidad de llegar a resultados más eficientes a través de procedimientos que incluyan los conceptos anteriormente mencionados.

La integración de hardware y código COTS, la aplicación de nuevos entornos de desarrollo, el cambio de los conceptos rectores clásicos por otros más flexibles con la incorporación de algunas nociones que pertenecen a la metodología Agile.

Estos cambios que se proponen requerirán de formas de pensamiento diferentes. Poner al cliente, ya sea interno o externo, en el centro del interés y hacerlo formar parte del proceso de desarrollo. No pensar a nivel de bit, sino con la estrategia del armado de un rompecabezas donde cada pieza cumpla una función específica.

Esta metodología de trabajo cambia el foco y requiere re especificar muchos preceptos, por ejemplo, el concepto de requisito, calidad, satisfacción, alcance, integración, interfaz, poniéndolos al servicio de procedimientos orientados a la satisfacción continua y a la colaboración estrecha con el usuario.

Qué sucedería, por ejemplo, si se definiesen en el transcurso del desarrollo de un sistema embebido, zonas permeables a cambios de especificaciones y zonas que no permitiesen modificaciones durante el proyecto.

En lugar de analizar la función de cada componente electrónico, estudiar cómo satisfacer

un requisito o realizar una tarea con módulos que pueden adquirirse en el mercado sin necesidad de diseño electrónico profundo, poniendo énfasis en la integración.

En resumen, se propone estudiar nuevos caminos de diseño electrónico a través del aporte de la integración, las nuevas plataformas de desarrollo, las facilidades actuales para generar prototipos y producciones de baja escala, incorporando conceptos de metodologías ágiles.

El objetivo principal es generar conocimiento sobre diseño de sistemas analógico-digitales a través del desarrollo de diseño e implementación de dispositivos de aplicación específica. Se busca que las nociones obtenidas en este proceso se capitalicen en los integrantes del equipo de trabajo en nuevas visiones que amplíen y diversifiquen las alternativas de solución a problemas tecnológicos.

Los objetivos secundarios son generar procedimientos superadores que modernicen y potencien los alcances de los desarrollos de sistemas embebidos, formar recursos humanos con alta capacitación para emprender y participar del mercado de diseño electrónico y optimizar los tiempos y procesos para generar soluciones tecnológicas a problemas reales que acerquen a la Universidad como centro generador de conocimiento, a las necesidades de la industria

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo de este proyecto está formado por ingenieros electrónicos y un especialista en seguridad informática. Como se mencionó anteriormente, este trabajo se desarrolló en el marco del proyecto de investigación: “Desarrollo de Dispositivos Electrónicos de Aplicación Específica a Baja Escala”.

El desarrollo del proyecto de investigación generó varias líneas de trabajo, de múltiples disciplinas. Dada la complejidad, fue necesaria la colaboración de varios expertos con amplia experiencia en la industria y la investigación académica.

Uno de los miembros del equipo de investigación se encuentra desarrollando su trabajo de tesis de posgrado de la Maestría en Informática de la UNLaM titulada: “ μ Framework: marco de referencia para desarrollos de sistemas

embebidos” y su tutor es el Mg. Jorge Eterovic, director del proyecto de investigación [10].

5. BIBLIOGRAFIA

- [1] Pedraza Suarez; Jaime Enrique Rodriguez Beltran; Edwin Alejandro. Reingeniería de firmware de controladora accesscan v. 4.0. <http://repositorio.uan.edu.co/handle/123456789/2287>. Bogotá, Colombia. 2020.
- [2] Amaro Calderón; Sarah Dámaris Valverde Rebaza; Jorge Carlos. Metodologías Ágiles. Trujillo, Perú. 2007.
- [3] Jean Paul Subra; Aurélien Vannieuwenhuyse. Scrum: un método ágil para sus proyectos. Ediciones ENI. ISBN 978-2-409-01292-1. Barcelona, España. 2018.
- [4] Tosini, Marcelo Alejandro; Todorovich, Elías; Vázquez, Martín Osvaldo et all. Metodologías de Diseño para Sistemas Embebidos. XV Workshop de Investigadores en Ciencias de la Computación. Universidad Autónoma de Entre Ríos. 2013.
- [5] Pablo Andrés García. Sistemas Embebidos de Tiempo Real con aplicaciones en Bioingeniería. La Plata, Argentina. 2019.
- [6] Louis Frenzel, Practical Electronic Design for Experimenters. Ed. McGraw Hill. ISBN: 978-1260456158. 2020.
- [7] Colin Walls. Embedded RTOS Design. Ed. Newnes. ISBN: 9780128228524. 2018.
- [8] Martin Tate. Off-The-Shelf IT Solutions - A practitioner's guide to selection and procurement. Ed. BCS Learning & Development Limited. ISBN: 9781780172583. 2018.
- [9]. Lucas Cordeiro; Raimundo Barreto et all. Agile Development Methodology for Embedded Systems: A Platform-Based Design Approach. 14th Annual IEEE International Conference and Workshops on the Engineering of Computer-Based Systems (ECBS'07). ISBN:0-7695-2772-8. Tucson, AZ, USA. 2007.
- [10] Fourcade A., Eterovic J., Pérez A., Rodofile G., μ Framework: marco de referencia para desarrollo de sistemas embebidos; CoNaIISI 2019; RIISIC-CONFEDI-UNLaM, San Justo. 2019.

ASI – Agentes y Sistemas Inteligentes

Extensiones a DEHIA en el marco de una arquitectura distribuida con microservicios

Jose Arcidiacono ¹, Alejandra B. Lliteras ^{2,3}, Patricia Bazán ¹

¹ UNLP, Facultad de Informática, LINTI ² UNLP, Facultad de Informática, LIFIA, ³ CICPBA

jarcidiacono@linti.unlp.edu.ar, alejandra.lliteras@lifa.info.unlp.edu.ar,
pbaz@info.unlp.edu.ar

RESUMEN

DEHIA es una plataforma para la creación y ejecución de actividades de recolección de datos que requieren de la intervención humana para realizar este objetivo.

La plataforma está definida sobre una arquitectura distribuida, basada en microservicios y ya cuenta con revisiones en torno a la distribución por componentes en contenedores independientes que aportan a la interoperabilidad pero que aumenta la complejidad operacional y requiere automatización y monitoreo.

En este trabajo, se presentan dos líneas de trabajo que extienden la plataforma en torno a la recolección de datos: 1- realizar la recolección ex-situ y 2- definir un componente de visualización de los datos recolectados.

Palabras clave: Microservicios, Recolección de datos, Interoperabilidad

CONTEXTO

Este proyecto surge como extensión del proyecto “DEHIA: Una plataforma liviana para definir y ejecutar actividades con intervención humana basada en workflows” [Arcidiacono, 2020], en el cual se presenta una aplicación web que permite a usuarios finales definir actividades que requieran de

la intervención humana con el fin de realizar recolección y análisis de datos, así como una aplicación móvil para ejecutar dichas actividades.

Dado que esta plataforma pone el foco en el análisis de los datos, además de la recolección de los mismos, resulta de interés poder completar parte de la arquitectura de microservicios en la que se basa DEHIA con un visualizador que permita un análisis integral, eficaz y flexible, siguiendo la estructura y filosofía de DEHIA.

Por otra parte, DEHIA realiza recolección de datos mediante una aplicación móvil, que permite obtener los datos in situ y tanto de manera online como offline (para enviar los datos recolectados cuando se logre conectividad online).

Así, un cliente web que permitiera recolectar datos ex-situ (es decir fuera del lugar donde se recolectaron) para DEHIA, daría la posibilidad de resolver los workflows de forma online y sin necesidad de tener instalada la aplicación de forma local. Aunque limita a que los usuarios tomen datos en un medio alternativo (por ejemplo, planillas de papel u anotaciones libres) y luego los carguen ya no in-situ.

1- INTRODUCCIÓN

Las arquitecturas orientadas a servicios (SOA - Services Oriented Architecture) tienen como objetivo la reutilización de componentes funcionales autónomos que puedan interoperar, componerse, integrarse y escalar para resolver funcionalidades más complejas.

DEHIA adopta una arquitectura basada en *micro*-servicios, un estilo de arquitectura que simplifica los contratos entre dichos servicios favoreciendo su composición donde cada uno de ellos ofrece una interfaz que permite la integración de nuevos componentes.

Dentro de la arquitectura de DEHIA hay tres componentes que representan la extensión planteada en este trabajo y son: el proveedor de identidad y autorización, el recolector de datos y el gestor de workflows.

DEHIA permite recolectar datos desde su aplicación móvil, en la cual los usuarios ejecutan los workflows de actividades mientras recorren de manera in-situ un lugar. Sin embargo, es posible que ciertos dominios requieran que los datos no fueran recolectados desde una aplicación móvil, o bien que los usuarios que recolectan los datos no quieran dejar prácticas arraigadas y bien probadas como la recolección de datos en planillas de papel. En tal caso, el foco principal, es lograr que los datos sean recolectados de alguna manera y posteriormente cargados de forma homogénea para su posterior uso. Para este caso en particular lo propuesto hasta el momento en DEHIA es insuficiente.

Adicionalmente, una vez que los datos son recolectados y almacenados, surge la necesidad de que los mismos puedan ser presentados de manera comprensible para las personas. Desde esta perspectiva, la

visualización de datos mediante gráficos ayuda a la comunicación y permite revelar información que los modelos y simulaciones pueden perder [Unwin, 2020], por ejemplo, valores atípicos y patrones locales [Unwin, 2020]. La visualización propuesta en DEHIA no incluye actualmente visualización de los datos mediante gráficos: sólo propone la visualización de datos en forma de tabla.

Se sabe que en la actualidad contar con datos es de suma importancia para diferentes acciones, por ejemplo, compararlos, detectar tendencias y mostrar información [Faronius Hofmann & Håkansson, 2021].

Hasta el momento se cuenta con una herramienta de recolección de datos (por parte de personas) y un mecanismo de almacenamiento apropiado para el uso que se le da actualmente a la plataforma. Sin embargo, sería interesante considerar un uso más extendido que llevaría al desafío de capturar, almacenar, manipular y mostrar grandes volúmenes de datos.

2- LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La extensión a DEHIA tiene por objetivo profundizar algunas de las líneas de investigación que la plataforma abarca. En particular, la extensión presentada en este trabajo se enfoca en las áreas, como 1- Integración de Componentes Distribuidos, 2- Interacción Humano- Computador, 3 - Usabilidad y Accesibilidad tanto web como móvil, 4- Recolección de datos 5- Visualización.

A partir del proyecto de investigación presentado en [Lamoureux, Z et al., 2019], los autores proponen criterios/aspectos a tener en cuenta a la hora de analizar

plataformas de Ciencia Ciudadana para la recolección de datos y realizan un benchmarking de algunas aplicaciones en dicho dominio, como por ejemplo, Zooniverse [Simpson, R. et al., 2014] y [Cox, J. et al., 2015] así como EpiCollect [Aanensen, D. M. et al., 2009] y [Gohil, R. et al., 2020] se observa la necesidad de proponer este cliente web ex-situ, con la intención de extender el alcance a más usuarios al momento de recolectar datos en DEHIA.

Por otra parte, la visualización de los datos recolectados y el conocimiento generado por las personas, permite traducir los datos en gráficos y elementos visuales que facilitan a los usuarios finales identificar patrones, extraer conocimiento significativo y distinguir las perspectivas que dicha información presenta. Así es como, a través de una adecuada visualización, el usuario final encuentra una manera accesible de interactuar con los datos y generar un posterior análisis, característica muy importante para construir una comunicación eficiente y concreta de los resultados obtenidos y el valor de los mismos. Lograr dichas visualizaciones puede resultar una tarea difícil para aquellas personas sin conocimientos específicos de este proceso y, particularmente, de las herramientas tecnológicas disponibles, es por eso que este componente de visualización se propone acortar la brecha entre la representación de los datos y sus usuarios finales.

La Figura 1 muestra la integración entre la plataforma DEHIA actual y las diferentes componentes relativas a cada línea de investigación, diferenciando en la misma aquellas líneas de trabajo que se encuentran en progreso de otras posibles.

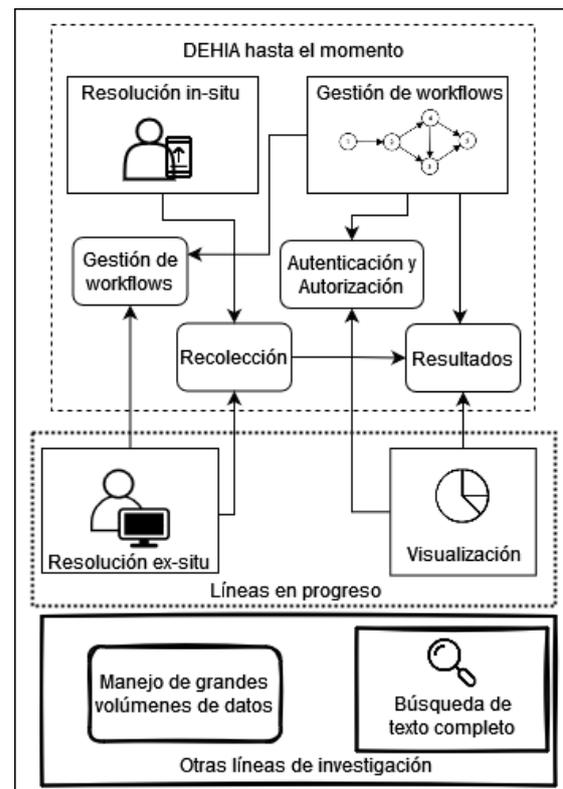


Figura 1 - Componentes de DEHIA a lo largo de las líneas de investigación

En la Figura 1 pueden observarse los componentes cliente – en este caso los que tienen interfaz de usuario – como rectángulos con línea llena, mientras que los servicios se encuadran en rectángulos con los bordes redondeados. Las flechas indican tanto la dependencia entre los componentes como la comunicación que sucede entre ellos a través de HTTP.

3- RESULTADOS ESPERADOS/OBTENIDOS

Dentro de los resultados esperados se encuentra el desarrollo e integración a la arquitectura actual de microservicios de DEHIA de un componente de visualización que permita el enriquecimiento de los datos recolectados mediante la plataforma y el análisis de los mismos a través de una variedad de gráficos y representaciones. Uno de los objetivos de este desarrollo en

particular es facilitar a usuarios finales el acceso a la información producto del proceso de recolección de datos con intervención humana, por lo cual se dio importancia a la facilidad de uso de la aplicación.

Finalmente, se espera lograr también otra extensión en forma de sitio web para la resolución ex-situ de las actividades creadas desde DEHIA, que esté integrado a la arquitectura y permita el envío de los datos cargados desde el mismo hacia la plataforma.

4- FORMACIÓN DE RECURSOS HUMANOS

La diversidad de líneas de investigación enunciadas, así como la gran variedad de aplicación de los resultados esperados, propone un foco de investigación claro y concreto para la formación de recursos humanos, que encuentran en esta propuesta diversas aristas de trabajo.

La propuesta aborda temas de arquitecturas distribuidas, ingeniería de software, HCI, accesibilidad y usabilidad web, así como aspectos relacionados a herramientas de autor, test cases, workflow con intervención humana, recolección, y visualización de datos.

Adicionalmente, la extensión vinculada con la visualización de datos será aplicada a un caso de estudio con perspectiva de género a partir del trabajo conjunto con la Prosecretaría de Derechos Humanos de la Universidad Nacional de La Plata.¹

El presente trabajo describe la línea de investigación que ha dado lugar a la formulación de propuestas de tesinas de grado por cada una de las extensiones, así como otra propuesta de tesina futura en

torno a la integración de un nuevo componente que incorpora un motor de búsqueda de texto completo.

5- BIBLIOGRAFÍA

[Aanensen, D. M. et al., 2009] Aanensen, D. M., Huntley, D. M., Feil, E. J., al-Own, F. A., & Spratt, B. G. (2009). EpiCollect: linking smartphones to web applications for epidemiology, ecology and community data collection. *PloS one*, 4(9), e6968.

[Arcidiacono, 2020] Arcidiacono, J. (2020). DEHIA: una plataforma liviana para definir y ejecutar actividades con intervención humana basadas en workflows. Tesina de Grado. Facultad de Informática, UNLP.

[Cox, J. et al., 2015] Cox, J., Oh, E. Y., Simmons, B., Lintott, C., Masters, K., Greenhill, A., ... & Holmes, K. (2015). Defining and measuring success in online citizen science: A case study of Zooniverse projects. *Computing in Science & Engineering*, 17(4), 28-41.

[Faronius Hofmann & Håkansson, 2021] Therese Faronius Hofmann and Linda Håkansson. (2021). Visualization Design Effects on Credibility and Data Perception, and the Importance of Digital Interaction. Uppsala universitet

[Gohil, R. et al., 2020] Gohil, R., Sharma, S., Sachdeva, S., Gupta, S., & Dhillon, M. S. (2020). EpiCollect 5: A Free, Fully Customizable Mobile-Based Application for Data Collection in Clinical Research.

[Lamoureux, Z et al., 2019] Lamoureux, Z., & Fast, V. (2019). The tools of citizen science: An evaluation of map-based crowdsourcing platforms. *Spatial Knowledge and Information Canada*, 7(4), 1.

¹ <https://unlp.edu.ar/ddhh>

[Simpson, R. et al., 2014] Simpson, R., Page, K. R., & De Roure, D. (2014, April). Zooniverse: observing the world's largest citizen science platform. In Proceedings of the 23rd international conference on world wide web (pp. 1049-1054).

[Unwin, 2020] Unwin, A. (2020). Why is Data Visualization Important? What is Important in Data Visualization?. Harvard Data Science Review, 2(1).

Aplicaciones de Machine Learning para el uso Sustentable de Recursos Naturales

Pablo Ezequiel Inchausti, Ana Martínez Saucedo, Leonardo Javier Amet, Pedro Martin Blanco, Guillermo Manuel Nievas y Luana Giusto

Universidad Argentina de la Empresa, Instituto de Tecnología (UADE - INTEC), CABA, Argentina
{pinchausti, anmartinez, lamet, peblanco, gunievas, luagiusto}@uade.edu.ar

RESUMEN

Se propone una investigación para predecir la ocurrencia de incendios forestales basada en el entrenamiento de Modelos de Machine Learning. Se utiliza para el entrenamiento de los modelos, datos de registros históricos provistos por las propias Asociaciones de Bomberos Voluntarios, datos del Servicio Meteorológico Nacional (SMN) e imágenes satelitales provistas por la NASA. Se propone extender la solución para abarcar el monitoreo de áreas en riesgo, mediante dispositivos de IoT en puntos fijos o móviles, y equipados con sensores y cámaras. El procesamiento de las imágenes se propone realizar mediante algoritmos de reconocimiento de imágenes para enviar alertas de posibles focos de incendios.

Palabras Clave: Machine Learning, IoT, Medio Ambiente, Incendios Forestales, Sustentabilidad

CONTEXTO

La investigación se desarrolla dentro del Instituto de Tecnología (INTEC) dependiente de la Universidad Argentina de la Empresa (UADE). Se inicia en el año 2021 como una Actividad de Ciencia y Técnica (ACyT).

El proyecto lleva el nombre “*Aplicaciones de Machine Learning para mejorar el uso de Recursos Naturales*” y se lo identifica con el código A21T03. Se propone una duración de dos años en el período 2021 y 2023 y se designa como investigador responsable a Mg. Pablo Ezequiel Inchausti.

Durante el año 2021 la investigación avanza por medio del desarrollo del Proyecto Final de Ingeniería (PFI) de la Ing. Ana Martínez

Saucedo, denominado “*AQUA: Prevención de Incendios Forestales En Pinamar con Machine Learning*”. El trabajo aporta a la investigación los fundamentos teóricos acerca del estudio de los incendios forestales y se desarrolla el primer prototipo funcional, denominado “*AQUA*”.

Respecto a la Ing. Ana Martínez Saucedo, en diciembre 2021 recibe una beca cofinanciada por UADE-CONICET para iniciar sus estudios doctorales.

Durante el año 2022 la investigación se extiende para incorporar la integración con dispositivos de IoT (en inglés, *Internet of Things*) que permita realizar acciones de alertas y monitoreo. Para el desarrollo de la línea de extensión sobre IoT, se integra al proyecto el Dr. Ing. Leonardo Javier Amet como especialista en Electrónica y Telecomunicaciones.

1. INTRODUCCIÓN

En Argentina han sido pocos los antecedentes de investigaciones y estudios para tratar la problemática de los incendios forestales. Y a su vez, los efectos producidos por el fuego han sido devastadores para el medio ambiente, con daños irreparables en la biodiversidad y pérdidas económicas millonarias que repercuten en la sociedad en su conjunto.

Un claro ejemplo de sus consecuencias son los incendios forestales que se produjeron en febrero de 2022 en la provincia de Corrientes, Argentina. De acuerdo a los informes divulgados por el INTA, el Instituto Nacional de Tecnología Agropecuaria, los incendios forestales arrasaron con 935.000 hectáreas, es decir, un equivalente al 11% de la superficie provincial [1].

En cuanto a la tendencia, se observa que la severidad de los incendios forestales es cada vez

mayor y así también la superficie final quemada. Y entre las causas de esta tendencia creciente, se identifican los efectos derivados del Cambio Climático y también el impacto de nuestra sociedad y su avance negativo sobre el medio ambiente [2].

En términos numéricos, desde el INPE, el Instituto Nacional de Investigación Espacial de Brasil, se indica que en Argentina se alcanzó en el 2020 una cifra record de 74.113 focos activos de incendios, representando un incremento cercano al 150% en términos interanuales [3].

Respecto a los tipos de estudios de incendios forestales, van desde la inclusión de sensores especializados, hasta satélites que permiten detectar focos de incendios, monitorear los cambios en la distribución y salud de la vegetación, determinar la temperatura de la superficie y estudiar el cambio climático a escala planetaria. Como los datos de estos satélites son de libre acceso, es posible utilizarlos para el entrenamiento de modelos de Machine Learning. [4]

Los desarrollos de modelos de incendios forestales se han enfocado en:

- Predicción espacial y temporal de incendios.
- Detección de incendios forestales.
- Predicción de área quemada por los incendios.
- Detección del área quemada a causa de incendios.
- Simulación de propagación de fuego una vez iniciado el incendio.

Respecto a los antecedentes de Machine Learning para predecir incendios, se puede mencionar a Rodrigues et al., 2014 [5] que compara la precisión de predicción de distintos modelos. También existen aplicaciones móviles gratuitas que calculan la probabilidad de incendios forestales a partir de índices meteorológicos como el FWI (*Forest fire Weather Index*, por sus siglas en inglés). Y respecto a Cardenas et al. [6] a nivel nacional desarrolla un sistema de predicción de incendios forestales en la provincia de Córdoba, con técnicas de Machine Learning basadas en redes neuronales y máquinas de soportes vectoriales (o SVM, por sus siglas en inglés, *Support Vector Machines*).

Objetivo:

Describir cómo la tecnología de Machine Learning (ML) puede contribuir al uso sustentable de los recursos naturales, iniciando el estudio con la prevención de incendios forestales, y luego extrapolar los conocimientos a otros contextos, en donde los recursos naturales resulten clave para el desarrollo.

2. LINEAS DE INVESTIGACION Y DESARROLLO

En el presente proyecto, que se encuadra dentro de la investigación aplicada, se propone desarrollar soluciones basadas en tecnologías de Machine Learning, para contribuir a la predicción y prevención de incendios forestales. Se propone entrenar modelos para predecir la probabilidad de ocurrencia de incendios forestales y generar alertas en base al reconocimiento de imágenes obtenidas por dispositivos de monitoreo.

El modelo de ML propuesto en AQUA se enmarca en las áreas de predicción espacial-temporal de incendios forestales y el área quemada por los mismos.

Para el entrenamiento de los modelos, desde el equipo de investigación se estuvo trabajando en conjunto con la Asociación de Bomberos Voluntarios de Pinamar (ABVP), para la digitalización de los registros históricos desde el año 2015. Los datos recolectados se complementan con datos abiertos de las plataformas Open Data, imágenes satelitales provistas por la NASA, y datos del servicio Meteorológico Nacional (SMN) en estaciones de monitoreo cercanas a la zona de estudio.

Como parte del procesamiento, se relaciona el Índice Normalizado de Vegetación (NDVI por sus siglas en inglés *Normalized Difference Vegetation Index*) con una serie de otras variables ambientales. El índice NDVI refleja cuan saludable se encuentra la vegetación en un área, y se considera que mientras más seca se encuentre la vegetación, mayores son las probabilidades de que se genere un incendio. Con estos números junto con otros indicadores ambientales, se hace posible identificar las zonas que se encuentran en riesgo.

La presentación de los resultados se proporciona mediante el prototipo funcional de

la aplicación denominada AQUA, que en su primera versión presenta sobre un mapa de Pinamar, tanto los registros históricos como las predicciones de las zonas de riesgo de incendios realizadas por los modelos entrenados con ML.

Como líneas de extensión y posibilidades de integración, se propone presentar los datos utilizados para el entrenamiento en una plataforma especializada en la visualización de datos medio ambientales junto con su contexto de interpretación. Esta plataforma se denominada “*Glaciar*” y fue desarrollada en UADE en el contexto de un Proyecto de Investigación (PID) identificado con el código PI18T04 [7]

Respecto a la línea de extensión con los dispositivos de IoT, se propone realzar la recolección primaria de los datos con fines de generación de alertas y monitoreo.

Por ejemplo, un caso concreto de aplicación de IoT es el desarrollo de redes de sensores inalámbricos para monitoreo de parámetros ambientales, o detección de fuego. Cada sensor inalámbrico es un dispositivo electrónico digital con procesadores embebidos que se encargan de medir los parámetros necesarios, procesarlos y enviarlos a una plataforma de IoT, desplegada, por ejemplo, en Amazon Web Services (AWS) para realizar la captura de los datos.

3. RESULTADOS OBTENIDOS/ESPERADOS

La recolección primaria de los datos de los incendios forestales para el entrenamiento de los modelos de Machine Learning, se realizó trabajando en conjunto con el cuartel de bomberos de Pinamar, digitalizando cinco años de registros históricos.

En el proceso de digitalización se utilizó la versión de gestión administrativa del software “*AQUA*”, que se presentó en las 50 JAIIO por parte de Ana Martínez Saucedo [8] y constituye un antecedente para el presente proyecto.

Durante el año 2021, en el marco del presente proyecto, Ana Martínez Saucedo desarrolló su trabajo de PFI de Ingeniería en Informática “*AQUA: Prevención de Incendios Forestales En Pinamar con Machine Learning*” que entre sus resultados se presenta el software “*AQUA*” con los modelos entrenados para

predecir incendios forestales en la ciudad de Pinamar, utilizando los datos de los bomberos voluntarios, datos e índices meteorológicos de la zona e imágenes satelitales de la NASA.

De forma adicional, debido a la gravedad de los efectos de los incendios forestales, y en especial, los desastres producidos por los incendios forestales en Corrientes en febrero de 2022 [1] que contó con una amplia cobertura de los medios de prensa, se generó un interés por divulgar el trabajo que se viene realizado dentro del presente proyecto de investigación. Y en ese contexto, se les solicitó a los investigadores compartir en los medios los avances y resultados obtenidos al momento. Como producto del proceso de difusión, al momento se publicaron dos artículos de prensa. El primero en el portal institucional de UADE describiendo en términos generales al proyecto [9] y el segundo, en el portal de la emisora Frecuencia Zero FM, transcribiendo una entrevista realizada a Ana Martínez Saucedo como parte del equipo del investigación [10].

Durante el año 2022, el presente proyecto tiene como objetivo extender su solución para la integrar la solución de predicción de incendios con Machine Learning, con monitoreo con dispositivos de IoT, por ejemplo, sensores, cámaras, y drones, integrándose con algoritmos de reconocimiento de imágenes incendios para identificar alertas de fuego.

4. FORMACIÓN DE RECURSOS HUMANOS

Como parte de la formación de recursos humanos, se realizó una primera convocaría para alumnos dirigida a alumnos de la carrera ingenierías informática, y como resultado de esa primera convocatoria se sumaron al equipo cinco integrantes.

Está planificado el lanzamiento de una segunda convocatoria a estudiantes de los últimos años de las carreras Gestión Ambiental, Ingeniería Informática, Ciencia de Datos e Ingeniería Electrónica. El objetivo de la convocatoria es tener alumnos que puedan investigar sobre el dominio del tema, alumnos que puedan realizar la analítica de los datos y el entrenamiento de los modelos predictivos, y alumnos que puedan investigar sobre la

integración con los dispositivos de IoT que se aplicarán en el proyecto.

También se contempla que del proyecto de investigación se deriven temas a desarrollar en trabajos de Proyecto Final de Ingeniería (PFI) para los alumnos de último año de las carreras de Ingeniería en Informática e Ingeniería Electrónica.

Sobre la audiencia interna, que la constituye la Universidad Argentina de la Empresa (UADE), se proponen actividades de transferencia describiendo a los prototipos funcionales, y dirigidas hacia las siguientes materias de la Ingeniería en Informática:

- *Inteligencia Artificial*
- *Arquitectura de Aplicaciones*
- *Tecnología y Medio Ambiente*

Se tiene en cuenta que la tecnología aplicada en los prototipos funcionales genera valor hacia la formación de los alumnos, al estar basados en plataformas de código abierto, y utilizando productos y técnicas de software de actualidad en el mercado.

Sobre las tecnologías que se utilizan, se mencionan entre las principales a React, Angular, Mongo, Node.js, Docker, Python, Scikit-learn, Tensorflow, entre otras, que se integran con las prácticas de Integración y Despliegue Continuo, conocidas con las siglas de CI/CD, para desplegar en entornos de nubes públicas como AWS y Azure. Siendo estas tecnologías de referencia ampliamente difundidas dentro de la industria del software actual.

Sobre la plataforma de despliegue, se utiliza a AWS como la principal plataforma de nube pública para desplegar las aplicaciones en un entorno público. También se utilizan los servicios de Machine Learning de AWS para realizar el entrenamiento de los modelos. En ese contexto, se pretende que el equipo de investigación alcance una sólida formación en los servicios de AWS, utilizando, en parte, cursos de AWS específicos de Arquitectura de Aplicaciones en la Nube y Machine Learning. Los cursos se realizan en la plataforma *AWS Academy* que desde el año 2021 tiene un convenio activo con UADE. También se está en proceso de gestión de créditos de uso de la plataforma AWS para investigadores [11] para

disponer de una cuenta de AWS con crédito para ser utilizada dentro el proyecto de investigación, principalmente en servicios de despliegue y entrenamiento a escala de los modelos predictivos.

Sobre la difusión académica, se propone la participación en congresos nacionales e internacionales con publicaciones de los resultados derivados del proyecto. Algunos ejemplos de estos congresos lo constituyen las Jornadas Argentinas de Informática (JAIIO), el Congreso Argentino de Ciencias de la Computación (CACIC) y la Escuela de Ciencias Informáticas (ECI).

En cuanto al aporte social, el trabajo inicia con la colaboración de la Asociación de Bomberos Voluntarios de Pinamar (ABVP), y la digitalización de sus registros históricos de incendios forestales para ser utilizados como datos de entrada para el entrenamiento de los modelos. Pero los procedimientos y aplicativos son extrapolables a los distintos Cuarteles de Bomberos de nuestro país, permitiendo acercar estas herramientas y tecnologías a las distintas regiones del Argentina, y a su vez permitiendo enriquecer la base de datos que nutren los entrenamientos para generar mejores predicciones que se encuentren al servicio de la sociedad.

5. BIBLIOGRAFÍA

- [1] P. Escobar, “Informe técnico del INTA: Corrientes registró una superficie de 935 mil hectáreas afectadas por los incendios”, *Argentina Forestal*, 25 de febrero de 2022. <https://www.argentinaforestal.com/2022/02/25/informe-tecnico-del-inta/>
- [2] M. D. Flannigan, B. J. Stocks, y B. M. Wotton, “Climate change and forest fires”, *Sci. Total Environ.*, vol. 262, n.º 3, pp. 221-229, nov. 2000, ISSN 00489697. doi: 10.1016/S0048-9697(00)00524-6.
- [3] INPE, “Monitoramento dos Focos Ativos por País - Programa Queimadas”. https://queimadas.dgi.inpe.br/queimadas/portal-static/estatisticas_paises/
- [4] P. Jain, S. C. P. Coogan, S. G. Subramanian, M. Crowley, S. Taylor, y M. D. Flannigan, “A review of machine learning applications

in wildfire science and management”, *Environ. Rev.*, vol. 28, n.º 4, pp. 478-505, dic. 2020, ISSN 1181-8700, 1208-6053. doi: 10.1139/er-2020-0019.

- [5] M. Rodrigues y J. de la Riva, “An insight into machine-learning algorithms to model human-caused wildfire occurrence”, *Environ. Model. Softw.*, vol. 57, pp. 192-201, jul. 2014, ISSN 1364-8152. doi: 10.1016/j.envsoft.2014.03.003.
- [6] M. Cardenas *et al.*, “Sistema de Predicción de Incendios Forestales para la Provincia de Córdoba”, 2016.
- [7] P. Inchausti, P. Romanos, M. Barturen, y B. D. Rossi, “Glaciar: software de visualización de datos de recursos hídricos y atmosféricos”, XI Congreso de Agro Informática (CAI) - JAIIO 48 (Salta, 2019), 2019, ISSN 2525-0949.
- [8] A. C. Martínez Saucedo, F. Connell, B. Ríos, y J. Perrotta, “AQUA: Sistema de administración y gestión documental de emergencias para cuarteles de bomberos”, p. 9, oct. 2021. ISSN 2451-7534.
- [9] UADE, “Modelo UADE para la prevención de incendios forestales”, 4 de febrero de 2022.
<https://www.uade.edu.ar/noticias/modelo-uade-para-la-prevencion-de-incendios-forestales/>
- [10] L. Coro, “Proponen un modelo para prevenir incendios forestales”, *Frecuencia Zero FM*, 24 de febrero de 2022.
<https://frecuenciazero.ar/proponen-un-modelo-para-prevenir-incendios-forestales/>
- [11] Amazon Web Services, Inc., “AWS Cloud Credit for Research”.
<https://aws.amazon.com/es/government-education/research-and-technical-computing/cloud-credit-for-research/>

Generador de proyectos de Software con transferencia a la comunidad

Mg. Rodolfo Bertone¹ C.C Marcelo Taruschio² Lic. Martin Correa³

Facultad de Ciencias Exactas e Ingeniería. Universidad Católica de La Plata.

¹ rodolfo.bertone@ucalp.edu.ar ² mtarus2012@gmail.com ³ martin.correa@ucalp.edu.ar

Resumen

El concepto “Usina de Ideas” define un espacio virtual y/o físico donde se generan, evalúan, proponen, discuten, acrecientan ideas de toda la comunidad de la facultad y/o universidad. De esta forma se busca generar un entorno que sea gestor de proyectos prácticos, que consistan en una investigación, presentación, integración, publicación científica y transferencia a la comunidad en su faz más práctica.

Este proyecto tiene como finalidad primaria la gestión de alternativas de trabajo para el desarrollo profesional de los alumnos en el marco de sus tesis de grado. Además, el análisis, estudio y desarrollo de soluciones informáticas para una nuestra Universidad son parte de las actividades que como grupo incipiente de investigación debemos encarar.

En FACEI, desde 2017 se vienen generando este tipo de propuestas y las mismas se plasmaron en varios trabajos científicos que fueron publicados en congresos internacionales de la especialidad.

Palabras clave:

Gestión de proyectos de software. Automatización y Control. Internet de las cosas. E-Reciclado. Tratamiento de residuos informáticos peligrosos

Contexto

Esta línea de Investigación se formaliza en 2022 siendo uno de los primeros proyectos de Investigación en el área Informática y Sistemas que tiene la FACEI (Facultad de Ciencias Exactas e Ingeniería) de UCALP (Universidad Católica de La Plata). Hasta la definición de este proyecto toda la tarea de investigación se desarrolla ad-hoc a partir de gestión de docentes y alumnos de esta casa. La formalización viene como parte de un plan rector de nuestra Universidad donde se está volcando mayores recursos a la investigación en general y a la transferencia de conocimiento en particular.

Introducción

Este proyecto marco busca definir un entorno dentro de nuestra Universidad donde se genere un contexto de investigación basados principalmente en analizar y discutir soluciones concretas a problemas específicos. Por el contexto de nuestra carrera y nuestro cuerpo docente no es posible generar laboratorios o institutos de investigación como lo cuenta Universidades Nacionales.

Desde 2016 se han presentado problemas a los alumnos, cuyo desarrollo necesitó de investigación, análisis e implementación de soluciones concretas. En varios de estos casos se obtuvieron soluciones innovadoras que generaron trans-

ferencias a la comunidad. Además, los resultados fueron presentados, aprobados y expuestos en congresos de la especialidad.

Cabe mencionar algunos de los trabajos realizados. En 2016 se publicó un artículo que presentaba una experiencia realizada en el marco de un proyecto de transferencia desarrollado dentro de la cátedra Proyecto de Sistemas de la carrera Ingeniería en Sistemas de Información de la Universidad Católica Argentina de La Plata. El proyecto, denominado SAAC, Sistema Automático de Asistencia al Cultivo, describió un modelo experimental con una solución integral capaz de controlar de manera automática diferentes tipos de cultivos desde un dispositivo móvil Android. El resultado obtenido presentó una combinación eficaz de Hardware, Software e ideas innovadoras, que a partir de una parametrización sencilla permite controlar y monitorear la temperatura, humedad, nivel de fertilizante en tierra e iluminación del cultivo, y de esta forma actuar sobre la superficie cultivada introduciendo mayor humedad, luz, calor o fertilizantes (1)

El segundo proyecto publicado fue el denominado Sistema de Identificación de personal. Los adultos mayores constituyen uno de los grupos más vulnerables en cuanto a la posibilidad de verse afectados en hechos de inseguridad en sus hogares. Una de las modalidades utilizadas por delincuentes, es la de hacerse pasar por representantes o empleados de empresas prestadoras de servicios (luz, gas, teléfono, etc.) y, por ejemplo, con la excusa de tener que realizar un arreglo interno en el domicilio, ingresan al mismo y cometen un acto delictivo. Ante esta problemática, surge como una posible solución el Sistema de Identificación de Personal (SIP). Este sistema tiene por finalidad permitir a la persona que se encuentre en el domicilio constatar de forma simple,

rápida y a bajo costo, mediante una solución informática, la veracidad del envío de personal por parte de las empresas prestadoras de servicios, con el fin saber de qué forma actuar con la persona que está, “del otro lado de la puerta”. (2)

El tercer proyecto realizó estudios en el ámbito asociado a la salud. Ahí se observó que existe un alto grado de intercambio de información entre los diferentes actores involucrados. Los datos se encuentran altamente fragmentados y distribuidos en múltiples sistemas no integrados, imposibilitando la comunicación y el intercambio de información entre ellos. El trabajo realizado analizó, investigó, describió y documentó los distintos niveles de interoperabilidad de datos de salud (IO), mencionando sus beneficios y barreras de implementación. (3)

El último trabajo realizado y presentado a un congreso fue el proyecto de control de tránsito en una Smart City. Este trabajo surge a partir de la siguiente premisa ¿cómo resolver los problemas de tránsito de una ciudad como La Plata aplicando tecnología? La capital de la provincia de Buenos Aires se caracteriza por tener un tránsito que se puede denominar caótico. Se debe tener en cuenta que, sobre los problemas clásicos de una ciudad cualquiera, La Plata cuenta con múltiples diagonales que complican más el tránsito. Teniendo varios factores en cuenta, se analizaron y plantearon soluciones novedosas a problemas conocidos. Se aplican conceptos ligados a GPS, sensores y placas Arduino e Internet, así como la posibilidad de contar con información en línea de un número de usuarios considerables lo que permitiría organizar al tránsito a partir del uso de la tecnología. (4)

Líneas de Investigación. Objetivos.

Como se definió anteriormente el objetivo general del proyecto es definir proyectos con transferencia al medio que resuelvan problemas, donde el software sea el elemento constitutivo esencial (eventualmente se podrían incluir desarrollos específicos de hardware).

Los Objetivos Específicos son: 1) Establecer reuniones de brainstorming donde participen alumnos y docentes, 2) Generar casos de estudio donde se analice la viabilidad de los subproyectos presentados, 3) Presentar el subproyecto de acuerdo a la modalidad definida en la normativa FACEI para tesis de grado, 4) Desarrollar el sub-proyecto y generar informes de avances, informes técnicos, monografías y/o publicaciones para congresos.

La metodología de trabajo involucra diversas actividades, que pueden ser interdisciplinarias y serán desarrolladas durante 2022, estableciendo el plazo de proyecto en 10 meses a partir del primero de marzo. En función de los resultados obtenidos y de la necesidad de mantener en el tiempo este tipo de actividades como generador de ideas, el proyecto podrá ser presentado nuevamente en años posteriores.

Cada subproyecto deberá tener su planificación temporal, el plazo mínimo de desarrollo se establece en cuatro meses. Es posible que simultáneamente se tengan varios subproyectos en ejecución.

Cada subproyecto deberá ser presentado al director, el cual lo pondrá a consideración de los participantes para su aprobación. En caso de tratarse de proyectos definidos como Tesis de Grado, deberá contar con el aval del director de carreras de sistemas.

Desarrollo e Innovación

En el marco definido en el apartado anterior se mencionan tres subproyectos ya iniciados.

El primero, denominado E-Reciclado. Procesamiento de material informático en desuso, consiste en desarrollar una serie de actividades que permitan generar un protocolo de tratamiento de estos residuos. Este protocolo contemplará actividades que van desde la recepción de los diferentes elementos, su clasificación y la realización de convenios específicos con instituciones o empresas que se encarguen de su tratamiento final. Dentro de nuestra Facultad, además de la carrera de sistemas, se cursan carreras de Seguridad e Higiene y, además, de Ingeniería Ambiental. Entonces el desarrollo de este proyecto contiene una idea de trabajo interdisciplinaria, que permite formar a los alumnos involucrados de tres áreas de conocimiento diferente.

A partir del proyecto anterior y bajo el mismo análisis de desarrollo interdisciplinario, se plantea el tratamiento de residuos peligrosos como son las pilas y baterías de los equipos de cómputo. Bajo este proyecto se estudiarán alternativas de procesamiento a partir de las normativas vigentes.

El tercer subproyecto que se encuentra en vías de realización se denomina AIS signal Detector. En el ámbito portuario y de navegación, la tecnología de sistemas de identificación automática ha mejorado notablemente a la navegación como seguridad, eficiencia, protección del ambiente e identificación de buques. Un Sistema de Identificación Automática (AIS, por sus siglas en inglés) se utiliza en todos los buques, puertos, boyas y balizados. Cada puerto tiene asociado un conjunto de artefactos navales que sirven para realizar marcas específicas que ayudan a la navegación. El sistema que se está desarro-

llo permite analizar el recorrido de los buques y poder determinar si alguno de ellos por fallos, errores y omisiones, colisionan los artefactos disponibles. Es factible que acciones como estas ocurran, por lo tanto, contar con sistemas de monitoreo permiten detectar situaciones anómalas y demostrar el hecho ocurrido es muy importante en particular debido al alto costo de reemplazo de los elementos electrónicos dañados. El sistema una vez finalizado será utilizado en una primera instancia en el puerto de la Ciudad de La Plata.

Resultados y Objetivos

Como se trata de un proyecto que se encuentra en una fase inicial, no se pueden mencionar resultados obtenidos. Sin embargo, y como se definió en la introducción del presente, subproyectos similares ya obtuvieron resultados importantes y publicados en congresos.

Formación de Recursos Humanos

El director del proyecto es un docente de la Facultad y sus tareas de investigación se realizan como una actividad complementaria. Los demás integrantes son profesores algunos con reconocida experiencia en la investigación.

La formación de recursos humanos es una actividad fundamental del proyecto. Cada subproyecto generado tiene y tendrá la participación de alumnos en su fase de terminación de carrera.

Referencias

(1) S.A.A.C. - Sistema Automático de Asistencia al Cultivo. Di Luca Federico, Di Iasi Luis, Marengo Javier, Monguillot Matías, Negri Juan, Bertone Rodolfo. CACIC 2016, San Luis Octubre 2016

(2) SIP (Sistema de Identificación de Personal), una solución informática al servicio de la seguridad de nuestros adultos mayores. Augusto José Markic, Leandro Germán Herrera, Rodolfo Bertone. CACIC 2018 Tandil. Octubre 2018

(3) Interoperabilidad e Integración entre Sistemas de Información en Salud. Marcela Demichelis, Martin Correa, Rodolfo Bertone. CACIC 2018 Tandil. Octubre 2018

(4) Control de tránsito en una Smart City. Juan Pablo Murdolo, Rodolfo Bertone. CACIC 2021 Salta. Octubre 2021

Deep Learning para aplicaciones astronómicas, visión por computadora y sistemas médicos.

F. Ronchetti^{1,2,3} , F. Quiroga^{1,2,4} , G. Rios^{1,4} , P. Dal Bianco^{1,4} , I. Mindlin¹ , L. Lanzarini^{1,2} ,
A. Rosete⁵ , R. Gamen⁶ , Y. Aidelman^{6,7}, C. Escudero⁷, N. Pereyra¹, E. Rucci^{1,2,3} .

¹ Instituto de Investigación en Informática LIDI, Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina.*

² Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina

³ Comisión de Investigaciones Científicas de la Pcia. De Bs. As. (CICPBA)

⁴ Becario postgrado UNLP

⁵ Universidad Tecnológica de La Habana “José Antonio Echeverría” (CUJAE), La Habana, Cuba

⁶ Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, La Plata, Argentina

⁷ Instituto de Astrofísica de La Plata (IALP CONICET), La Plata, Argentina

* Centro asociado de la Comisión de Investigaciones Científicas de la Pcia. De Bs. As. (CIC)

Contacto: fronchetti@lidi.info.unlp.edu.ar

CONTEXTO

Esta presentación corresponde a las tareas de investigación que se llevan a cabo en el III-LIDI en el marco del proyecto F025 “Sistemas inteligentes. Aplicaciones en reconocimiento de patrones, minería de datos y big data” perteneciente al Programa de Incentivos (2018-2022).

RESUMEN

Esta línea de investigación se centra en el estudio y desarrollo de Sistemas Inteligentes para la resolución de problemas de reconocimiento de patrones en imágenes, video y datos médicos, utilizando técnicas de Aprendizaje Automático clásicas, junto con Redes Neuronales Convolucionales y Aprendizaje profundo. El trabajo presentado describe diferentes casos de aplicación en visión por computadora, Astronomía y predicción de diabetes.

Una de las líneas de investigación principales que se continúa desarrollando es el reconocimiento de lengua de señas. Este es un problema complejo y multidisciplinar, que presenta diversos subproblemas a resolver como el reconocimiento del intérprete, la segmentación de manos, la clasificación de diferentes configuraciones y de un gesto dinámico, entre otros.

En esta área se está estudiando la forma de reconocer formas de mano de la Lengua de Señas con conjuntos de datos de tamaño

reducido, dada la falta de datos de entrenamiento para este dominio. Además, se están utilizando Redes Recurrentes para reconocer señas dinámicas, utilizando la base de datos LSA64 de Lengua de Señas Argentina [2].

Por último, se están utilizando Redes Generativas Adversarias (GANs) para aumentar bases de datos de formas de mano, con el objetivo de complementar desde otro enfoque el entrenamiento de modelos para su clasificación.

Por otro lado, se está estudiando la forma en que las redes neuronales codifican la invarianza a las transformaciones y otras propiedades transformacionales, con el objetivo de poder analizar y comparar estos modelos. De esta forma se espera poder mejorar los modelos de clasificación de objetos transformados, en particular, de formas de mano.

Siguiendo con la línea de reconocimiento de patrones en imágenes, se está llevando a cabo una colaboración con investigadores de la Facultad de Astronomía y Geofísica de la UNLP para crear modelos de clasificación de imágenes de objetos celestes. Además, se está desarrollando un sistema para recuperar la información de placas.

Por último, se estudiaron modelos de Aprendizaje Automático para la predicción temprana de la enfermedad de diabetes.

Palabras clave: Redes Neuronales, Redes Convolucionales, Redes Recurrentes, Visión

por Computadoras, Lengua de Señas, Redes Generativas Adversarias, Invarianza, Equivarianza, Imágenes Astronómicas.

1. INTRODUCCION

El Instituto de Investigación en Informática LIDI (III-LIDI) tiene una larga trayectoria en el estudio, investigación y desarrollo de Sistemas Inteligentes basados en distintos métodos de Aprendizaje Automático y Redes Neuronales

Como resultado de estas investigaciones se han diseñado e implementado técnicas originales aplicables a la clasificación y el análisis de características de objetos en imágenes, generación de imágenes para aumentación de datos, y estudio del funcionamiento de las redes neuronales. En relación con esta línea, actualmente se están desarrollando los siguientes temas:

1.1. Reconocimiento Lengua de Señas

En esta línea de investigación, el objetivo es la clasificación de gestos en videos de Lengua de Señas. Por un lado, se está trabajando con la base de datos LSA64 [11]. Esta consiste en un registro de 64 señas de la Lengua de Señas Argentina. Se compararon diversas técnicas en el estado del arte del Aprendizaje Automático basadas en Redes Neuronales. Específicamente, se compararon tanto arquitecturas basadas en Redes Recurrentes y Convolucionales, como distintas estrategias de preprocesamiento para optimizar la calidad del reconocimiento. También se están analizando los modelos entrenados para comprender mejor el impacto de estas estrategias de preprocesamiento y de la forma de representación lograda por cada tipo de arquitectura [2][11]. Por otro lado, se está confeccionando una base de datos masiva y realista de la Lengua de Señas Argentina, procesando videos de YouTube de un canal especializado en la temática. Se espera generar una base de datos robusta con cientos de horas de señas continuas.

Adicionalmente, se están comenzando a utilizar Redes Generativas Adversarias (GANs) para generar imágenes artificiales

relacionadas con la lengua de señas. Este tipo de redes permitirá aumentar las bases de datos de formas de mano, con el objetivo de complementar desde otro enfoque el estudio de modelos y algoritmos de clasificación para bases de datos con pocos datos etiquetados. Estas investigaciones son llevadas a cabo en marco de una tesis doctoral financiada por la UNLP a través de una beca de postgrado.



Figura 1: Activaciones de las capas convolucionales de una Red Recurrente para reconocer señas en video.

1.2. Identificación de Personas con Riesgo de Diabetes y Prediabetes

La Diabetes Tipo 2 (DT2) es una enfermedad crónica caracterizada por una disminución precoz y progresiva de la masa y de la función de las células beta del páncreas. Debido a su creciente prevalencia en combinación con su elevado costo de atención, constituye un serio problema de salud pública, por lo que se han realizado grandes esfuerzos por desarrollar estrategias efectivas para su prevención y tratamiento, así como para evitar sus complicaciones crónicas. En ese sentido, resulta importante reconocer que las consecuencias negativas de esta enfermedad comienzan en una etapa previa conocida como prediabetes, la cual implica un riesgo elevado de desarrollar DT2 en los siguientes años.

El desarrollo de la DT2 es un proceso lento y progresivo condicionado por factores genéticos, ambientales y de comportamiento. Aunque no existe una cura definitiva para esta enfermedad, varios estudios han demostrado que se puede prevenir o demorar su aparición en personas con prediabetes a través de la adopción de un estilo de vida saludable y/o asociado con la ingesta de diversos fármacos. En Argentina, una iniciativa de este tipo es el

Programa PPDBA desarrollado por el CENEXA (CONICET-UNLP-CIC) [10].

La detección de DT2 y prediabetes representa un verdadero desafío para la medicina debido a la ausencia de síntomas patogenómicos y/o la falta de conocimiento de los factores de riesgo asociados. Es por eso que frecuentemente una persona pueda pasar meses (o incluso años) sin saber que se encuentra en riesgo.

Los modelos existentes para predicción de diabetes y prediabetes no necesariamente aplican a la población argentina [13] y, hasta donde llega nuestro conocimiento, no existe modelo ni herramienta similar disponible en nuestro medio que permita identificar personas con alta probabilidad de tener estas enfermedades. Brevemente, esta línea propone desarrollar y validar modelos predictivos de diabetes y prediabetes específicos para la población argentina utilizando técnicas de Aprendizaje Automático. Se cuenta con acceso a la base de datos del PPDBA y apoyo del equipo médico de CENEXA. La concreción de esta línea representaría un avance en el conocimiento y un instrumento útil para los sistemas de salud de Argentina (e incluso de la región).

1.3. Métricas de Equivarianza

Las redes neuronales son modelos tradicionalmente considerados como de caja negra. En los años recientes, se han realizado varios esfuerzos para comprender su funcionamiento de forma tal que el mismo sea más predecible o modulable.

La invarianza y equivarianza a las transformaciones son propiedades deseables en varios modelos de redes debido a que nos permiten razonar más fácilmente respecto a su funcionamiento.

En los últimos años, varios modelos fueron propuestos para añadir invarianza a la rotación y otras transformaciones en CNNs [3]. No obstante, no está claro como estos modelos impactan en el aprendizaje usual de los pesos de la red.

Por este motivo, se continua con la utilización de las métricas previamente definidas [4] para

caracterizar modelos de redes neuronales típicos, ya sea desde capas muy utilizadas como Batch Normalization, Dropout, Max-Pooling, arquitecturas completas como Residual Networks, VGG o AllConvolutional, y arquitecturas especializadas como TI-Pooling.

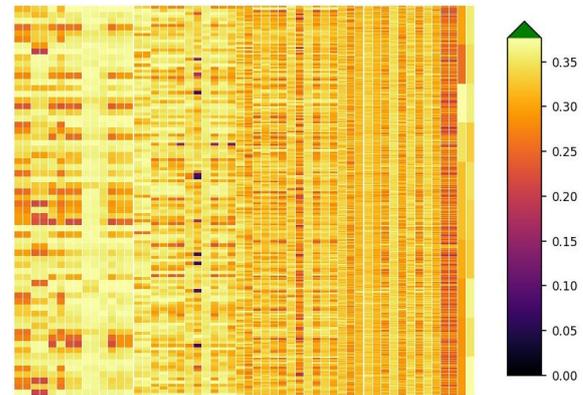


Figura 2. Invarianza por capas y unidades de una CNN con arquitectura ResNet.

1.4. Análisis de Imágenes Astronómicas

En los últimos años, la cantidad de información astronómica disponible se ha multiplicado de forma exponencial. En consecuencia, diversas tareas que anteriormente se realizaban de forma manual o semi-manual deben ahora automatizarse aún más.

En este ámbito, en el III-LIDI se están desarrollando dos proyectos. El primero consiste en determinar modelos de clasificación a partir de información fotométrica de estrellas Be [5][12]. Estas estrellas son estrellas no-supergigantes cuyo espectro exhibe emisión en línea H α . Para el análisis de muchos objetos astronómicos, los datos fotométricos son relativamente más fáciles de obtener debido al menor tiempo de uso del telescopio. Por lo tanto, existe una necesidad creciente de utilizar información fotométrica para identificar automáticamente objetos para estudios más detallados, especialmente estrellas con líneas de emisión H α como las estrellas Be. En este trabajo se evaluó el uso de redes neuronales para identificar candidatos a estrella Be a partir de un conjunto de estrellas tipo OB. Las redes se

entrenaron utilizando un subconjunto etiquetado de las bases de datos VPHAS + y 2MASS, con filtros u, g, r, H α , i, J, H y K. Para evitar el efecto del enrojecimiento, se propuso y se evaluó el uso de índices Q para mejorar la generalización del modelo a otras bases de datos. Para validar el enfoque, se etiquetó manualmente un subconjunto de la base de datos y se usó para evaluar modelos de identificación de candidatos. También se etiquetó un conjunto independiente de datos utilizando validación cruzada [7]. Los modelos entrenados lograron obtener un 25% de Sensibilidad (*Recall*) fijando la precisión (*Precision*) al 99%. Actualmente se está trabajando en agregar nuevas bases de datos de otros investigadores y así como otras áreas del cielo.

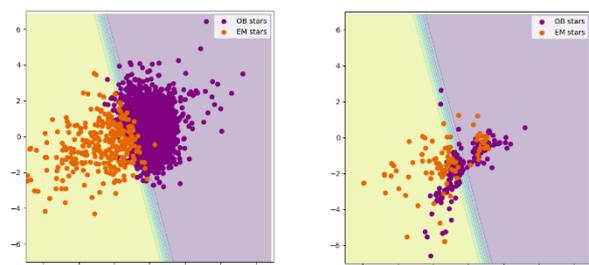


Figura 3. Reducción a 2 features para los conjuntos de datos de Mohr-Smith (izquierda) y Liu (derecha). Las líneas de decisión corresponden a una función sigmoidea de un regresor logístico entrenado en Mohr-Smith.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

- Redes neuronales profundas, convolucionales y recurrentes.
- Invarianza y auto-equivarianza en redes neuronales.
- Reconocimiento de lenguaje de señas.
- Generación de imágenes con GANs.
- Análisis de datos astronómicos.
- Identificación de Personas con Riesgo de Diabetes y Prediabetes.

3. RESULTADOS OBTENIDOS/ESPERADOS

- Comparación de modelos especializados para bases de datos con pocas muestras para la clasificación de formas de mano.
- Desarrollo de métricas de invarianza y auto-equivarianza para redes neuronales.
- Análisis de modelos para el reconocimiento de lengua de señas en video.
- Redes generativas para la creación de datos artificiales en la Lengua de Señas.
- Modelo de clasificación de estrellas Be generalizable a distintos conjuntos de datos.
- Sistema automático de recortes de espectros estelares.
- Modelos de predicción de diabetes y prediabetes.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo de la línea de I/D aquí presentada está formado por: 2 profesores con dedicación exclusiva, 1 investigador CIC-PBA, 2 becarios de posgrado de la UNLP con dedicación docente, 1 becario CIC, 1 becario CIN, 6 tesistas, 1 profesor extranjero y 3 investigadores externos.

Dentro de los temas involucrados en esta línea de investigación, en los últimos dos años se han finalizado 2 tesis de doctorado, 2 tesis de especialización, y 5 tesinas de grado de Licenciatura.

Actualmente se están desarrollando 2 tesis de doctorado, 1 tesis de especialista, 5 tesinas de grado de Licenciatura y 2 trabajos finales de Ingeniería en Computación. También participan en el desarrollo de las tareas becarios y pasantes del III-LIDI.

5. REFERENCIAS

- [1] Quiroga, F., Antonio, R., Ronchetti, R., Lanzarini, L., Rosete, A. A Study of Convolutional Architectures for Handshape Recognition applied to Sign Language, publicado en el XXIII Congreso Argentino de Ciencias de la Computación (CACIC 2017) (pp. 13-22). 2017
- [2] Cornejo Fandos, U., Rios, G., Ronchetti, F., Quiroga, F., Hasperué, W., Lanzarini, L. Recognizing Handshapes using Small Datasets, publicado en el XXV Congreso Argentino de Ciencias de la Computación (CACIC 2019, Rio Cuarto) (pp. 105-114). 2019.
- [3] Quiroga F., Ronchetti F., Lanzarini L., Fernandez-Bariviera A. Revisiting Data Augmentation for Rotational Invariance in Convolutional Neural Networks. International Conference on Modeling and Simulation in Engineering, Economics and Management (MS'2018 GIRONA). 2018.
- [4] Quiroga, F., Torrents-Barrena, J., Lanzarini, L., & Puig, D. Measuring (in) variances in Convolutional Networks. In Conference on Cloud Computing and Big Data (pp. 98-109). Springer, Cham. 2019.
- [5] Aidelman Y., Escudero C., Ronchetti F., Quiroga F., Lanzarini L. Reddening-Free Q Indices to Identify Be Star Candidates. Cloud Computing, Big Data & Emerging Topics. pp 111-123. Springer International Publishing. 2020.
- [6] Goodfellow I. J., Pouget-Abadie j., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative Adversarial Networks. NIPS'14 Proceedings of the 27th International Conference on Neural Information Processing Systems. v2. pp 2672-2680. 2014.
- [7] Jaschek M., Slettebak A., Jaschek C. *Be star terminology*. Be Star Newsletter. 1981.
- [8] Potash, P., Romanov, A., and Rumshisky, A. Ghostwriter: Using an lstm for automatic rap lyric generation. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 1919-1924. 2015.
- [9] Young, T., Hazarika, D., Poria, S., and Cambria, E. Recent trends in deep learning based natural language processing. *iee Computational intelligenCe magazine*, 13(3):55. 2018.
- [10] Gagliardino J. J., Etchegoyen G., Bourgeois M., Fantuzzi G., García S., González L., Elgart J. F., Ré M., Ricart A., Ricart J. P., Spinedi E., “Prevención primaria de diabetes tipo 2 en argentina: estudio piloto en la provincia de buenos aires,” *Revista Argentina de Endocrinología y Metabolismo*, vol. 53, no. 4, pp. 135 – 141, 2016.
- [11] Iván Mindlin, Facundo Quiroga, Franco Ronchetti, Pedro Dal Bianco, Gastón Ríos, Laura Lanzarini, Waldo Hasperué. “A Comparison of Neural Networks for Sign Language Recognition with LSA64”. JCC-BD&ET: Conference on Cloud Computing, Big Data & Emerging Topics. pp 104-117. Springer, Cham. Junio 2021.
- [12] Aidelman Y., Escudero C., Ronchetti F., Quiroga F., Granada A., Lanzarini L. “Identificación de candidatas a estrellas Be utilizando redes neuronales”. *Boletín de la Asociación Argentina de Astronomía*. Vol. 62. pp.62-64. 2021.
- [13] Choudhury A., Gupta D. A Survey on Medical Diagnosis of Diabetes Using Machine Learning Techniques. In: Recent Developments in Machine Learning and Data Analytics. Advances in Intelligent Systems and Computing, vol 740. Springer, Singapore. 2019.

DESARROLLO DE TÉCNICAS DE INTELIGENCIA COMPUTACIONAL PARA EL ANÁLISIS DE DATOS GENÓMICOS

Pablo Javier Vidal^{a,b}, Jessica Andrea Carballido^c, Ana Carolina Olivera^{a,b}, Matías Gabriel Rojas^a y Mariel Denise Volman Stern^b

^a*Instituto para las Tecnologías de la Información y las Comunicaciones. Consejo Nacional de Investigaciones Científicas y Técnicas. Universidad Nacional de Cuyo (ITIC-UNCuyo)*

^b*Facultad de Ingeniería - Universidad Nacional de Cuyo (FING-UNCuyo)*

^c*Instituto de Ciencias e Ingeniería de la Computación, Consejo Nacional de Investigaciones Científicas y Técnicas, Universidad Nacional del Sur (ICIC-CONICET-UNS)*

contacto: pjvidal@conicet.gov.ar, pablo.vidal@ingenieria.uncuyo.edu.ar

CONTEXTO

Esta presentación corresponde a las tareas de investigación que se llevan a cabo en el ITIC-UNCuyo en conjunto con la Facultad de Ingeniería de la UNCuyo y la Universidad Nacional del Sur en el marco del proyecto 06/B081-B (2019-2021) y una Beca Doctoral del CONICET.

RESUMEN

Esta línea de investigación se centra en el diseño y desarrollo de técnicas de Inteligencia Computacional en combinación con otros métodos con el propósito de contribuir al área de Ciencias de la Computación aplicando el conocimiento desarrollado a problemas de bioinformática, en particular aquellos de las áreas de genómica estructural.

Palabras clave: Bioinformática, Genómica Estructural, Inteligencia Computacional.

1 INTRODUCCIÓN

El conocimiento de las secuencias del genoma humano y de otras especies permite que se realicen investigaciones del genoma como un todo. Para analizar el conjunto de genes la bioinformática trabaja con diversas técnicas computacionales que están a disposición de los científicos debido a que pueden procesar e interpretar una gran cantidad de datos en n -dimensiones y reconstruir y separar estos conjuntos para tareas de clasificación o generar regresiones numéricas para tareas

de predicción. Esto permite determinar automáticamente si un caso de estudio presenta una patología específica o bien, si ciertas características del DNA (Deoxyribonucleic Acid) pueden permitir una clasificación de nuevos casos en forma rápida y eficiente, entre otras posibilidades. Nuestro grupo de investigación está abocado al diseño y aplicación de técnicas de Inteligencia Computacional para el análisis de datos genómicos está formado por investigadores del Instituto para las Tecnologías de la Información y las Comunicaciones de la Universidad Nacional de Cuyo y una Investigadora de la Universidad Nacional del Sur especialista en Bioinformática. Asimismo, participan del grupo una alumna de la Facultad de Ingeniería de la UNCuyo y un Becario Doctoral del CONICET.

Entre los problemas abordados por el grupo de investigación se encuentra el ensamblado de fragmentos de cadenas de DNA (Deoxyribonucleic Acid Fragment Assembly Problem, DNA-FAP). El cual consiste en encontrar dado un conjunto de cientos o miles de fragmentos de DNA, que pueden contener errores, la secuencia de DNA original a partir de las permutaciones de los fragmentos que mejor representen a dicha secuencia [Pev00].

Una vez secuenciada la cadena de DNA de interés, independientemente del objetivo, es indispensable compararla con las

secuencias disponibles en las diferentes bases de datos. Este análisis permite inferir algunas métricas de similitud entre las secuencias y llevarse a cabo diferentes tipos de análisis (filogenéticos, evaluación de la conservación de los dominios proteicos, las estructuras terciarias y secundarias, entre otros).

Existen dos tipos de alineamientos, el de pares de secuencias (Pairwise Sequence Alignment, PSA) y el alineamiento múltiple de secuencias (Multiple Sequence Alignment, MSA). El primero, consiste en la alineación de dos secuencias, mientras que el segundo implica alinear tres o más. El enfoque comúnmente adoptado para abordar el problema del MSA es el alineamiento progresivo que consiste en dividir el problema en subproblemas y resolverlos mediante PSA. Este enfoque es susceptible a quedar atascado en óptimos locales, debido a que una falla en la alineación de los primeros pares puede afectar a la solución final.

Nuestro grupo ha obtenido resultados promisorios en estos temas que se detallan en la Sección 3.

2 LINEAS DE INVESTIGACIÓN y DESARROLLO

- Diseño de novedosas técnicas de Inteligencia Computacional.
- Aplicación de Inteligencia Computacional a problemas de Bioinformática.

3 RESULTADOS OBTENIDOS/ESPERADOS

3.1 Avances en bioinformática.

En lo que respecta a secuenciamiento, alineamiento y ensamblado de fragmentos de cadenas de DNA se ha abordado el problema del secuenciamiento de las cadenas de DNA. Se diseñó un algoritmo híbrido basado con Cuckoo Search para el

secuenciamiento de cadenas de DNA [RCOV20a]. En este trabajo, se híbrida al *Cuckoo Search* con dos búsquedas locales diferentes para mejorar las capacidades de búsqueda del algoritmo canónico. Se lleva a cabo una evaluación numérica de ambas propuestas y la versión canónica utilizando un conjunto de datos de referencia bien conocido. Los resultados demuestran que la hibridación mejora la búsqueda y transforma la búsqueda del cuco en un procedimiento robusto con el potencial de tratar también con secuencias más largas o secuencias de longitud desconocida. Considerando el problema de MSA el grupo ha evaluado la capacidad de un Algoritmo Genético Celular Memético (Memetic Cellular Genetic Algorithm, MCGA) para realizar el alineamiento múltiple de secuencias [RCOV20c]. Se propone un algoritmo basado en el CGA combinado con un algoritmo de búsqueda local basada en la inserción, eliminación y reubicación de espacios en la secuencia. Con esto, se busca integrar la lenta difusión de la mejor solución y la rápida convergencia a un óptimo del CGA con la capacidad de identificar a los mejores vecinos de cada solución que posee el algoritmo de búsqueda local.

Asimismo, se ha trabajado con la caracterización multi-objetivo de la selección de características para microarrays de datos de cáncer y su impacto en las soluciones [DPOV20]. Se ha diseñado un algoritmo híbrido que combina el algoritmo genético celular (*Cellular Genetic Algorithm*, CGA) con una búsqueda de vecindario variable (*Variable Neighborhood Search*, VNS) diseñada para este problema. Esta técnica se comparó con otros métodos del estado del arte y los resultados experimentales indicaron que nuestra propuesta supera numéricamente a las otras evaluadas [ROCV20].

3.2 Avances teóricos en Inteligencia Computacional

La capacidad de precisión de una técnica de aprendizaje automático debe ser lo suficientemente robusta ante la aparición de diversos muestras o instancias en diferentes tipos de problemas. Cada una de estas técnicas presentan parámetros ajustables que mejoran o empeoran la capacidad de clasificación de los mismos [EMS19]. La construcción de un modelo de aprendizaje automático eficiente es un proceso complejo y lento que implica determinar el algoritmo apropiado y obtener una arquitectura de modelo óptima mediante el ajuste de sus hiperparámetros [KJ+13]. El ajuste incorrecto lleva una precisión inexacta y computacionalmente más costosa. Teniendo en cuenta que el ajuste de los hiperparámetros de una técnica de ML (*Machine Learning*) es un problema de optimización difícil, se pueden utilizar métodos de inteligencia computacional como enfoque para su resolución. Con respecto al ajuste de hiperparámetros para un mejor desempeño de un SVM (*Support Vector Machine*), el grupo ha comenzado a explorar modificaciones a las configuraciones por defecto utilizadas usualmente en el SVM. En [RCOV20b] se propuso la utilización de meta-heurísticas para la optimización de los hiperparámetros del SVM utilizando un *kernel Wavelet*. Se observó que todas las meta-heurísticas lograron alcanzar valores altos de precisión en comparación a la configuración por defecto del SVM utilizada comúnmente en la literatura. Por otro lado, el algoritmo genético celular utilizado se destacó en el tiempo de ejecución promedio que le demandó alcanzar el criterio de parada, lo que sugiere una mayor eficiencia con respecto a las otras propuestas. De la misma forma, se evidenció que el *kernel Wavelet* es capaz de lograr una distribución adecuada de los datos, siempre y cuando posea una correcta configuración de sus parámetros.

El caso de estudio utilizado fue un conjunto de datos relacionados a la enfermedad retinopatía diabética.

4 FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo está formado por: Un Profesor Adjunto con Dedicación Simple de la Facultad de Ingeniería de la UNCuyo (Investigador Asistente del CONICET), una Profesora Adjunta con Dedicación Exclusiva de la Universidad Nacional del Sur (Investigadora Adjunta del CONICET), Una Profesora Titular con Dedicación Semiexclusiva de la Facultad de Ingeniería de la UNCuyo (Investigadora Adjunta del CONICET), una Becaria EVC-CIN de la UNCuyo y un Becario doctoral del CONICET

El Ing. Rojas es dirigido por el Dr. Vidal y la Dra. Carballido en su doctorado que se encuentra realizando en la Facultad de Ciencias Exactas y Naturales de la UNCuyo en el tema “*Desarrollo de Metaheurísticas Aplicadas a Genómica Funcional y Estructural*”.

La Srta. Volman Stern es dirigida por los Dres. Vidal y Olivera en su Beca EVC-CIN en el tema *procesamiento de imágenes utilizando inteligencia artificial* [VOV21].

5 BIBLIOGRAFIA

- [DPOV20] Dussaut, J., Ponzoni, I., Olivera, A., y Vidal, P. Algoritmos evolutivos multiobjetivo aplicados a la selección de características en microarrays de datos de cáncer. *Entre Ciencia e Ingeniería* 14, 28 (dic. 2020),40–45, 10.31908/19098367.2014.
- [EMS19] Elshawi, R., Maher, M., y Sakr, S. Automated machine learning: State-of-the-art and open challenges. arXiv preprint arXiv:1906.02287 (2019)
- [KJ+13] Max Kuhn, Kjell Johnson, et al. *Applied predictive modeling*, volume 26. Springer, 2013.

- [Pev00] Pevzner, P. Computational Molecular Biology: An Algorithmic Approach. MIT Press, 2000.
- [POV19] Patrana, S., Olivera, A. C., y Vidal, P. J. Análisis del algoritmo genético celular para el problema de ensamblado de cadenas de ADN. *Informes Científicos Técnicos - UNPA* 3, 11 (2019), 236–248, 10.22305/ict-unpa.v11.n3.804.
- [RCOV20a] Rojas, M. G., Carballido, J. A., Olivera, A. C., y Vidal, P. J. Hybrid cuckoo search for solving DNA fragment assembly problem. In *IV Congreso Internacional de Ciencias de la Computación y Sistemas de Información* (2020).
- [RCOV20b] Rojas, M. G., Carballido, J. A., Olivera, A. C., y Vidal, P. J. Optimización de support vector machine mediante metaheurísticas para clasificación de retinopatía diabética. In *Simposio Argentino de Inteligencia Artificial de JAIIO* (2020).
- [RCOV20c] Rojas, M. G., Carballido, J. A., Olivera, A. C., y Vidal, P. J. A memetic cellular genetic algorithm for multiple sequence alignment. In *Proceedings of the 2020 IEEE Biennial Congress of Argentina* (November 2020).
- [ROCV20] Matías Gabriel Rojas, Ana Carolina Olivera, Jessica Andrea Carballido, and Pablo Javier Vidal. A memetic cellular genetic algorithm for cancer data microarray feature selection. *IEEE Latin America Transactions*, 2020.
- [VOV21] Volman Stern, M. D., Olivera, A. C., y Vidal, P. J. Paralelización del filtro convolución para imágenes digitales. In *Anales del V Congreso Internacional de Ciencias de la Computación y Sistemas de Información* (2021).

INTELIGENCIA COMPUTACIONAL APLICADA A MOVILIDAD URBANA

Ana Carolina Olivera^{a,b}, Pablo Javier Vidal^{a,b}, Enrique Gabriel Baquela^c, Marisa Haderne^b, Mariel Denise Volman Stern^b y Gastón Cavallo^b

^a*Instituto para las Tecnologías de la Información y las Comunicaciones. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Universidad Nacional de Cuyo (ITIC-UNCuyo)*

^b*Facultad de Ingeniería. Universidad Nacional de Cuyo (FING-UNCuyo)*

^c*Regional San Nicolás, Universidad Tecnológica Nacional (UTN-RSN)*

contacto: acolivera@conicet.gov.ar, carolina.olivera@ingenieria.uncuyo.edu.ar

CONTEXTO

Esta presentación corresponde a las tareas de investigación que se llevan a cabo en el ITIC-UNCuyo en conjunto con la Facultad de Ingeniería de la UNCuyo en el marco del proyecto 06/B081-B titulado *Diseño y Desarrollo de Herramientas Metaheurísticas y Aplicaciones Novedosas Orientadas a Smart Mobility* (2019-2021) financiado por la UNCuyo y el PICT 2020-SERIEA-00743 titulado *Diseño y desarrollo de estrategias novedosas aplicadas a problemas de tráfico vehicular y transporte público de pasajeros urbano* financiado por la Agencia Nacional de Promoción de la Investigación, el desarrollo Tecnológico y la Innovación, Argentina.

RESUMEN

Esta línea de investigación se centra en diseñar novedosas metodologías basadas en Inteligencia Computacional o hibridar existentes a fin de abordar satisfactoriamente problemas de Movilidad Urbana como los asociados al transporte público urbano y el tráfico vehicular. El trabajo presentado describe la línea de investigación que estamos llevando adelante y los resultados que se han obtenido hasta el momento.

Palabras clave: Inteligencia Computacional, Movilidad Urbana, Optimización.

1. INTRODUCCIÓN

Nuestro grupo de investigación está formado por investigadores del Instituto para las Tecnologías de la Información y las Comunicaciones de la Universidad

Nacional de Cuyo con amplia trayectoria en investigación y desarrollo en técnicas de Inteligencia Computacional aplicadas a distintos problemas complejos. Asimismo, participan del grupo docentes y alumnos de la Facultad de Ingeniería de la UNCuyo y de la Universidad Tecnológica Nacional Regional San Nicolás.

Movilidad urbana comprende una variedad de problemas derivados del movimiento de las personas a través de diferentes medios públicos o privados, motorizados o no [Ced21]. Existen diferentes medios de transporte insertos en las redes de tráfico de una ciudad. Este proyecto se enfoca en tráfico y transporte público urbano.

2. LINEAS DE INVESTIGACIÓN y DESARROLLO

- Diseño de novedosas hibridaciones de técnicas de Inteligencia Computacional.
- Aplicación de Inteligencia Computacional a problemas del área de Movilidad Urbana.

3. RESULTADOS OBTENIDOS/ESPERADOS

En lo que respecta a las líneas de investigación se realizaron diversos desarrollos en el campo de la inteligencia computacional. A continuación, se muestra un resumen de cada una.

3.1 Movilidad Urbana: Tráfico Urbano

El grupo de investigación realizó una

aproximación mono-objetivo de la optimización de los ciclos de los semáforos en la cual las únicas variables a optimizar que se consideran son aquellas del flujo vehicular. Nuestra motivación devino de contemplar que con sólo dichas variables se pueden obtener buenos resultados que impacten de manera positiva sobre la disminución de la polución (sea combustible fósil o no) y el gasto de combustible sin contar con estas en la función a optimizar puesto que dependen del parque automotor. Como caso de estudio utilizamos dos instancias reales, con información precisa del tráfico, el microcentro de la ciudad de Posadas en Argentina y el barrio Jacinto Vera en Montevideo, Uruguay. Para la instancia de Posadas, de las estrategias testeadas, el algoritmo de evolución diferencial (ED) fue el que obtuvo mejores resultados, alcanzando reducciones del 30% en tiempo de espera de los vehículos y hasta un incremento del 4% en la velocidad promedio de los vehículos. En lo que respecta a factores de contaminación, el algoritmo ED consiguió emisiones bajas y estables sin picos de contaminación en comparación con lo obtenido con la configuración actual de los semáforos. Asimismo, el algoritmo ED consigue una reducción significativa en el combustible consumido aun cuando la velocidad de los vehículos se incrementó. Esto demuestra que nuestra estrategia consigue mejorar la fluidez de los vehículos sin impactar en su consumo. Los resultados se encuentran publicados en la IEEE Latin America Transactions [VO19]. En ese mismo sentido se evalúa una aproximación similar para el microcentro de Mendoza [VO20] con resultados satisfactorios.

3.2 Movilidad Urbana: Transporte Público de Pasajeros

En lo que respecta a transporte público de pasajeros en un trabajo previo realizamos una combinación con el problema del

tráfico urbano considerando la optimización de los ciclos de las luces para el sector de la Avenida Garzón (Montevideo, Uruguay) que contiene un corredor rápido de buses (BRT, *Bus Rapid Transit*) mejorando la fluidez de los buses a través de un algoritmo evolutivo paralelo multi-objetivo. Los resultados muestran que el algoritmo evolutivo propuesto logra una mejor calidad de servicio en comparación con el estado actual, mejorando hasta un 15.3% la velocidad promedio del autobús y un 24.8% la velocidad promedio de otros vehículos. El trabajo se encuentra publicado en [NMA+19].

3.3 Distribución de vacunas en países en vías de desarrollo

Considerando que en 2020 sobrevino de forma abrupta la pandemia derivada del SAR-COV-2 nos abocamos al problema que poseen muchos países en vías de desarrollo que no accedían en forma automática a grandes cantidades de vacunas. En este sentido, se propuso un modelo de distribución de la vacuna COVID-19 y se probaron varias técnicas de inteligencia computacional. La propuesta se testeó en el escenario argentino. Se obtuvo una importante demostración del impacto de una distribución optimizada de la vacuna en territorios como el de Argentina [BO22].

3.4 Diseño de nuevas técnicas de Inteligencia Computacional

Un inconveniente cuando se aborda la resolución de problemas reales es el tiempo computacional asociado a la evaluación de las soluciones obtenidas. Durante nuestras investigaciones desarrollamos un modelo híbrido que combina *Kriging* (predicción) con algoritmos multi-objetivo en particular el conocido Non-Dominated Sorting Genetic Algorithm II [BO19] obteniendo excelentes resultados.

Se avanzó también en el estudio filtros para imágenes satelitales que sirven como información de partida para problemas relacionados como movilidad urbana [VOV21].

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo está formado por: Una Profesora Titular con Dedicación Semiexclusiva de la Facultad de Ingeniería de la UNCuyo (Investigadora Adjunta del CONICET), Un Profesor Adjunto con Dedicación Simple de la Facultad de Ingeniería de la UNCuyo (Investigador Asistente del CONICET), un Profesor Titular de la Universidad Tecnológica Nacional Regional San Nicolás, una Profesora Titular con dedicación Semiexclusiva de la Facultad de Ingeniería de la UNCuyo, una Becaria EVC-CIN de la UNCuyo y un alumno de la Licenciatura en Ciencias de la Computación de la UNCuyo.

Dentro de los temas abarcados por esta línea de investigación el Dr. Baquela finalizó su Doctorado en Ingeniería en el 2020 bajo la dirección de la Dra. Olivera. Se está desarrollando una Beca EVC-CIN en *procesamiento de imágenes utilizando inteligencia artificial* que puede utilizarse no solo para movilidad urbana sino también en otras áreas.

5. BIBLIOGRAFIA

[BO19] Baquela, E. y Olivera, A. A novel hybrid multi-objective metamodel-based evolutionary optimization algorithm. *Operations Research Perspectives* 6 Elsevier (2019).
 [BO22] Baquela, G. E. y Olivera, A. C. *Humanitarian Logistics in the Disaster Risk Reduction Perspective*. Springer, 2022, chapter Optimising Distribution of Limited COVID-19 Vaccines: Analysing Impact in Argentine, pp. 273–291. ISBN 978-3-030-90877-5. 10.1007/978-3-030-

90877-5_9.

[Ced21] Ceder, A. A. Urban mobility and public transport: future perspectives and review. *International Journal of Urban Sciences* 25, 4 (oct 2021), 455–479, 10.1080/12265934.2020.1799846.

[NMA+19] Nesmachnow, S., Massobrio R., Arreche, E., Mumford, C., Olivera, A. C., Vidal, P. J. y Tchernykh, A. Traffic lights synchronization for Bus Rapid Transit using a parallel evolutionary algorithm. *International Journal of Transportation Science and Technology* 8, 1 (2019), 53–67.

[VO19] Vidal P. J. y Olivera, A. C. Management of urban traffic flow based on traffic lights scheduling optimization. *IEEE Latin America Transactions* 17, 1 (jan 2019), 102–110.

[VO20] Vidal P. J. y Olivera, A. C.. Avoided traffic jams optimising the scheduling of signalised intersection of Mendoza downtown and surrounding. In *IV Congreso Internacional de Ciencias de la Computación y Sistemas de Información* (November 2020).

[VOV21] Volman Stern, M. D., Olivera, A. C., y Vidal, P. J. Paralelización del filtro convolución para imágenes digitales. In *Anales del V Congreso Internacional de Ciencias de la Computación y Sistemas de Información* (2021).

APLICACIÓN DE UNA ESTRATEGIA Y DE TÉCNICAS DE INTELIGENCIA Y ANALÍTICA DE NEGOCIO A LOS SISTEMAS DE INFORMACIÓN DEL MINISTERIO DE SALUD DE LA PROVINCIA DE SAN JUAN

Aguilera, C.¹, Romagnano, M.^{1,2}

¹Instituto de Informática, FCEFN, Universidad Nacional de San Juan

²Departamento de Informática, FCEFN, Universidad Nacional de San Juan

cynthiaguilerap@gmail.com, maritaroma@iinfo.unsj.edu.ar

RESUMEN

En la actualidad, en todos los niveles organizacionales, los datos, la información y el conocimiento se han convertido en uno de los recursos más valiosos para la toma de decisiones. Ante una inquietud o consulta de cualquier tipo, se debería acceder rápidamente a grandes volúmenes de datos. Sin embargo, su almacenamiento, procesamiento y posterior análisis, representan uno de los problemas más críticos a resolver.

El Ministerio de Salud de la provincia de San Juan, debido al gran volumen de datos que maneja y dado que los mismos se encuentran dispersos en distintos sistemas, ha comenzado a desplegar un sistema de salud integral. En este sistema confluirán todos los actores y sistemas de la salud provincial, tanto públicos como privados. Se espera que en el corto plazo un gran volumen de información, referida a la salud de los sanjuaninos, esté digitalizada. Es decir, que se cuente con un cúmulo importante de valiosa información, accesible a pacientes, profesionales y centros asistenciales. Pero, además, que estos sistemas de información contribuyan, principalmente, a la elaboración de políticas públicas del sector. Sin embargo, aun contando con estos sistemas de información integrados y digitalizados si no se dispone de recursos humanos y tecnológicos que permitan transformar esos datos en información y ésta en conocimiento,

difícilmente se pueden hacer inferencias, pronosticar y en ciertos casos prescribir acciones a seguir.

Por lo tanto, la presente investigación consistirá en definir una estrategia a seguir que, aplicando una determinada técnica y herramienta tecnológica de BI y BA, permita la visualización y análisis de los datos de los sistemas de información del Ministerio de Salud de la provincia de San Juan. Posteriormente, adquirir patrones de comportamiento, generar predicciones y posibles escenarios de acción (prescripciones).

Palabras clave: inteligencia de negocios, analítica de negocio, sistema de información, salud, gestión, herramientas.

CONTEXO

El presente trabajo se encuentra enmarcado en el proyecto “Aporte de los Sistemas de Inteligencia de Negocios a los Sistemas de Información Organizacionales para la Toma de Decisiones”, presentado en la convocatoria del Consejo de Investigaciones Científicas y Técnicas y de Creación Artística (CICITCA) de la Universidad Nacional de San Juan, para ser desarrollado durante el período comprendido entre 01/01/2020 al 31/12/2022.

Las tareas de investigación se desarrollan en el Laboratorio de Sistemas de Información, en el ámbito del Instituto de Informática de la

Facultad de Ciencias Exactas, Físicas y Naturales, UNSJ.

El equipo de trabajo viene aplicando técnicas y herramientas provistas por la inteligencia y la analítica del negocio a los sistemas de información del Ministerio de Salud provincial. Se espera aportar con visualizaciones, análisis y patrones de comportamiento, realizando predicciones que contribuyan a tomar certeras decisiones, a nivel ministerial y/o estatal.

1. INTRODUCCIÓN

A medida que crece la información, la exigencia que surge en las organizaciones para administrarla, se convierte en una necesidad y a la vez en un problema de cómo obtener la validez adecuada de los datos, razón por la cual el análisis de datos, su administración y la inteligencia de negocio es la prioridad tecnológica de hoy (Rodríguez Merchan, Martínez Guzmán, 2018).

El análisis de negocios se ha convertido en una frase de moda en la era actual de la economía de la experiencia. Principalmente, la proliferación de Internet y las tecnologías de la información han hecho de la analítica empresarial un área de aplicación sólida. Por otro lado, no se puede negar su impacto significativo en los campos de las tecnologías de la información, los métodos cuantitativos y las ciencias de la decisión (Cegielski y Jones-Farmer 2016).

Estas tendencias se vienen aplicando en varios dominios de aplicación. En cuanto a salud, los autores Fontana, Del Villar, Musanti, Schmidt, Bigatti y Moine (2020) afirman que la cantidad de información existente en los sistemas transaccionales utilizados en salud pública es abrumadora y analizarla es una tarea compleja. Más allá de los reportes convencionales, existen técnicas avanzadas de explotación de datos, que permiten organizar y estudiar la información almacenada desde diversas perspectivas. Para ello, se basan en la implementación de técnicas de Inteligencia de Negocios para unificar toda la información de interés en un Data Warehouse, a fin de obtener análisis

estadísticos mediante herramientas de explotación de datos que posibiliten la toma de decisiones basada en datos reales.

Por otro lado, Ali-Hassan (2019) realizó una investigación que se basó en presentar el entorno sanitario y la necesidad de análisis de datos e inteligencia empresarial en el ámbito de salud. Describió la diferencia entre los datos y la información y como ambos juegan un papel importante en la toma de decisiones, empleando un conjunto de herramientas analíticas que puedan describir eventos que han sucedido en el pasado, proporcionando un diagnóstico médico.

Actualmente, los profesionales de la salud, tienen acceso a grandes cantidades de datos en forma de registros del personal, registros electrónicos de pacientes, hallazgos clínicos, diagnósticos, medicamentos recetados, diagnóstico por imágenes, salud móvil, recursos disponibles, etc. Gestionar los datos y analizarlos para comprenderlos correctamente y utilizarlos para tomar decisiones es un desafío para los gerentes y los profesionales de la salud. Además, las herramientas de análisis de datos, también conocidas como análisis de negocios o herramientas de inteligencia, se están volviendo más poderosas, más accesibles y más fáciles de usar (Ali-Hassan, 2019).

La Inteligencia de Negocios (BI) y la analítica, en conjunto con abundante cantidad de datos y tecnologías avanzadas, han ganado recientemente una gran popularidad debido a su capacidad para mejorar el rendimiento en cualquier industria o campo de aplicación. La analítica, considerada por muchos como parte de BI, extrae, manipula y analiza datos, transformándolos en información que ayuda a los profesionales a tomar decisiones bien informadas. Apoya la acción y la generación de conocimiento. En el campo de la salud, la analítica jugará un papel importante en áreas como el diagnóstico, las admisiones y la prevención (Ali-Hassan, 2019).

Por otro lado, Justin Heinze describe en su artículo: “A medida que la inteligencia empresarial se convirtió en una frase comúnmente conocida a fines de la década de 1990 y principios de la de 2000, decenas de

nuevos proveedores llegaron al mercado. Durante ese período, hubo dos funciones básicas de BI: producir datos e informes, y organizarlos y visualizarlos de manera presentable. Sin embargo, quedaban dos problemas importantes que frenaban esta fase de desarrollo de la tecnología: la complejidad y el tiempo”.

El problema de las primeras herramientas de BI era que eran poco intuitivas y difíciles de usar. Además, para generar informes y acceder a la información, un usuario sin conocimiento específico debía recurrir a los servicios del departamento de TI. A finales de los 90 y principios del año 2000 fue cuando realmente se empezó a entender el potencial del uso de los softwares de Business Intelligence. El mercado empezó a florecer y proliferaron los fabricantes y proveedores de este tipo de herramientas. Además, estas soluciones empezaron a mejorar, haciéndose cada vez más usables para cualquier profesional, de forma que sin tener conocimientos informáticos se podía acceder, recopilar y analizar la información sin tener que recurrir al departamento de TI (Media, 2021).

Media afirma que el reto que surgió en ese momento fue el de combinar soluciones de BI rápidas y fáciles de utilizar por usuarios no técnicos, y que al mismo tiempo fuesen capaces de proporcionar información segura y confiable (Media, 2021).

Una herramienta para saber en qué punto de innovación y nivel de desarrollo están las empresas dedicadas a la tecnología en el mercado a nivel mundial, es el Cuadrante Mágico Gartner (Figura 1). Esta herramienta permite a las empresas que contratan servicios y soluciones en TI tener una visión de conjunto de una determinada área de productos o servicios tecnológicos, y poder tomar las mejores decisiones en sus procesos de transformación digital (ISC sd, 2019).

Este cuadrante nos ayuda a saber qué posición ocupan los proveedores y en qué punto de desarrollo se encuentran en cuanto a visión de mercado y poder de implementación, siendo definido para cada sector TI de acuerdo a dos criterios clave: la

amplitud de la visión y la capacidad para la ejecución. Estos dos criterios se convierten en los ejes del gráfico, los cuales generan 4 zonas diferenciadas de calificación para definir a las empresas de TI, las cuales son:

- Los aspirantes (Challengers). Según la web de Gartner: “tienen una buena ejecución del negocio actualmente y son capaces de dominar un gran segmento de mercado, pero aún no demuestran un real entendimiento de la dirección en que va el mercado”.
- Los líderes (Leaders). Desarrollan bien su negocio de acuerdo a una adecuada visión actual del mercado y están bien posicionados para el futuro.
- Los jugadores de nichos específicos (Niche Players). Se enfocan con éxito en un segmento de mercado específico, pero muchos de ellos no adquieren en su desempeño una visión global y no se caracterizan por hacer grandes innovaciones o por superar a sus competidores.
- Los visionarios (Visionaires) Entienden hacia dónde van los mercados e incluso pueden tener una idea para cambiar las reglas y paradigmas, pero aún no son capaces de llevar a cabo estas ideas por completo o con éxito.



Figura 1: Cuadrante Mágico de Gartner para Plataformas Analíticas y de Business Intelligence 2021. Fuente: Gartner, Inc.

El descubrimiento de datos, el cual

solía estar limitado a la pericia de especialistas de analíticas avanzadas, ahora es algo que todos pueden hacer usando herramientas de BI. Y no solo eso, estas herramientas dan las perspectivas que se necesitan para lograr cosas como el crecimiento, resolver problemas que sean urgentes, recolectar datos en un sitio, predecir resultados futuros y muchísimo más (Haije, 2019).

En cuanto a un Sistema de Información (IS), Peña los define como un conjunto de elementos interrelacionados con el propósito de prestar atención a las demandas de información de una organización, para elevar el nivel de conocimientos que permitan un mejor apoyo a la toma de decisiones y desarrollo de acciones (Peña, 2006). Cohen y Asín Lares (2005) lo definieron como un conjunto de elementos que interactúan entre sí con el fin de apoyar las actividades de una empresa o negocio. En un sentido amplio, un sistema de información no necesariamente incluye equipos electrónicos (hardware). Sin embargo, en la práctica se utiliza como sinónimo de “sistema de información computarizado”.

Los Sistemas de Información en Salud (SIS) constituyen el núcleo principal de soporte para la toma de decisiones en las instituciones sanitarias. Dentro de los SIS, los indicadores de salud (IS) reflejan numéricamente los acontecimientos medidos en el continuum salud-enfermedad. El sistema integrado de información en salud tiene por objetivo normalizar, integrar y organizar toda la información en salud disponible en sus sistemas de información, en un repositorio accesible y seguro, así como la distribución de la información de la forma más conveniente, para facilitar la toma de decisiones (Solera, Labordena, Loyola, 2010).

Generado por la pandemia que azota el mundo, la salud, es uno de los sectores que entró en una de las mayores fases de cambio, por lo tanto, a través de este trabajo de investigación se vislumbra la importancia de crear una solución de inteligencia de negocios, permitiendo que existan

alternativas de cómo modernizar la parte administrativa-tecnológica-social. De esta manera, generar procesos más rápidos, ágiles, que permitan también respuestas efectivas por parte de los expertos de salud. Asimismo, la implementación de este tipo de soluciones permite que en el sector público se puedan optimizar recursos lo que llevaría a recortes en gasto de papelería, formularios, bodega, entre otros sectores (Palacios, Medina, Ochoa, Mireya y Torres, 2020).

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

La presente investigación se basará en analizar y experimentar con un sistema de salud integral, propuesto por el Ministerio de Salud de Neuquén en acuerdo con el Gobierno de San Juan; denominado ANDES.

En consecuencia, se está trabajando en la definición de una estrategia a seguir que, aplicando una determinada técnica y herramienta tecnológica de BI y BA, permita la visualización y análisis de los datos de los sistemas de información del Ministerio de Salud de la provincia de San Juan. Posteriormente, se prevé adquirir patrones de comportamiento, generar predicciones y posibles escenarios de acción (prescripciones).

Para llevar a cabo el estudio, se está realizando una investigación de tipo exploratoria-descriptiva-experimental, usando fuentes de datos primarias y secundarias.

Para ello, se están desarrollando pruebas y análisis con la base de datos donde se encuentran implementados los sistemas de información (Figuras 2 y 3).

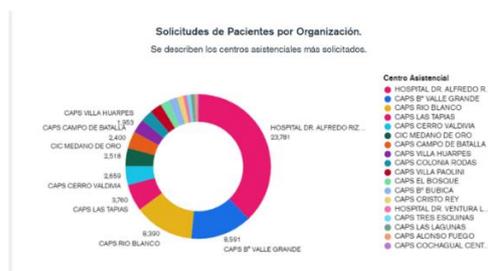


Figura 2: Visualización en MongoDB Charts, donde se aprecian solicitudes de pacientes por cada organización registrada. Fuente: Elaboración propia.

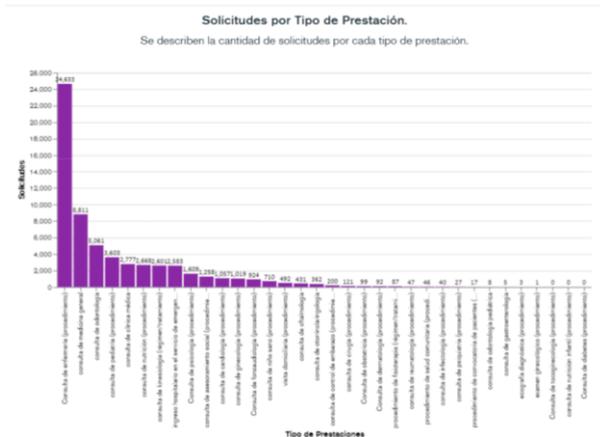


Figura 3: Visualización en MongoDB Charts, donde se aprecian solicitudes por tipo de prestación. Fuente: Elaboración propia.

Posteriormente, se procederá a la identificación de los beneficios derivados, al integrar técnicas y una aplicación de BI y BA a los sistemas de información del Ministerio de Salud provincial. Luego, se analizarán e informarán los resultados.

Finalmente, se establecerá la estrategia a seguir. Además, se procederá a la difusión de los resultados preliminares y a la documentación de los resultados.

3. OBJETIVOS ESPERADOS

Como objetivo se espera definir una estrategia a seguir que, sistematizando visualizaciones, análisis y predicciones, contribuyan con conocimiento a certeras y propicias tomas de decisiones, a nivel ministerial y/o gubernamental.

4. FORMACIÓN DE RECURSOS HUMANOS.

Este trabajo se encuentra enmarcado dentro de la temática planteada para desarrollar la tesis de la carrera de Licenciatura en Sistemas de Información, de la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de San Juan, San Juan.

Contribuye a su desarrollo el Instituto de Informática, donde la alumna Aguilera realiza sus tareas de investigación, bajo la dirección de la Mg. Lic. Romagnano.

5. BIBLIOGRAFÍA

- ALI-HASSAN, H. (2019). Healthcare data analytics and BI.
- CEGIELSKI, C. G., y JONES-FARMER, L. A. (2016). Knowledge, skills, and abilities for entry-level business analytics positions: A multi-method study. *Decision Sciences Journal of Innovative Education*, 4(1), 91–118.
- COHEN, K., Y ASÍN LARES, E. (2005). Sistemas de información para los negocios: un enfoque de toma de decisiones.
- FONTANA, L., DEL VILLAR, R., MUSANTI, J., SCHMIDT, N., BIGATTI, C., MOINE, J. M. (2020). Inteligencia de negocios y salud: análisis de emergencias médicas por accidentes en la vía pública. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/115780>
- HAWLEY, D. (2016). Implementing business analytics within the supply chain: success. *The Electronic Journal Information Systems Evaluation*, 19(2), 112–120.
- HEINZE, J. (2020). History of Business Intelligence. Disponible en: <https://www.betterbuys.com/bi/history-of-business-intelligence/>
- LÓPEZ BENÍTEZ, Y. (2018). *Business Intelligence. ADGG102PO. IC Editorial. 1º Edición. Andalucía – España. Vol. 1, 151 pag. Año 2018. ISBN: 978-84-9198-467-2.*
- MARTINEZ GUZMAN, E. M., RODRIGUEZ MERCHAN L. V. (2018). Análisis de Herramientas del Enfoque de Inteligencia de Negocios: Caso de Estudio Datos de la Banca Corporativa y de Inversión. Colombia. Disponible en: <http://polux.unipiloto.edu.co:8080/00004523.pdf>
- MEDIA (2021). ¿Qué es Business Intelligence (BI) y qué herramientas existen? Disponible en: <https://blog.signaturit.com/es/que-es-business-intelligence-bi-y-que-herramientas-existen>
- MEDINA, E. H., OCHOA, J. D., PALACIOS J. A., TORRES M.M. (2020). *Business Intelligence aplicado al sector Salud. Disponible en: https://www.researchgate.net/publication/346844881_Business_Intelligence_aplicado_al_sector_Salud*
- PEIRO R. (2020). Sistema de Información. Disponible en: <https://economipedia.com/definiciones/sistema-de-informacion.html>
- PEÑA AYALA, A. (2006). *Ingeniería de software: Una guía para crear sistemas de información. México DF (México): Instituto Politécnico Nacional.*
- ROLDÁN SALGUEIRO, J. L., CEPEDA-CARRIÓN, G., Y GALÁN GONZÁLEZ, J. L. (2012). Los sistemas de inteligencia de negocio como soporte a los procesos de toma de decisiones en las organizaciones - *Papeles de Economía Española*, No 132, 2012. ISSN: 0210-9107.
- SUN, Z., ZOU, H., Y STRANG, K. (2015). Big Data Analytics as a Service for Business Intelligence. Disponible en: https://www.researchgate.net/publication/282867145_Big_Data_Analytics_as_a_Service_for_Business_Intelligence, junio de 2022.

Optimización de la logística de distribución utilizando técnicas de la Inteligencia Artificial

Gabriela Minetti, Carolina Salto, Hugo Alfonso, Carlos Bermúdez,
M.Juliana Dielschneider Del Bono, Javier Vargas
Laboratorio de Investigación en Sistemas Inteligentes (LISI)
Facultad de Ingeniería - Universidad Nacional de La Pampa
Calle 110 Esq. 9 (6360) General Pico - La Pampa - Rep. Argentina
Te. / Fax: (02302) 422780/422372, Int. 6302
e-mail: {saltoc, minetti, alfonsoh, bermudezc, juliana, jvargas}@ing.unlpam.edu.ar

Resumen

La logística en Argentina está influida por el decisivo peso de la producción primaria en la organización del sistema, por una orientación prevalente hacia las exportaciones globales y las dirigidas al mercado regional sudamericano, y por una matriz de transporte dominada por el transporte automotor de cargas y la concentración portuaria. La logística remite a flujos de materiales y de información; a lugares de manipulación, depósito y transformación de las mercancías; a redes y nodos de circulación; y a tiempos de movimiento y no movimiento que responden a aspectos materiales (las infraestructuras, los transportes y las cargas) y también a aspectos funcionales (los servicios, las normativas y regulaciones). En suma, la logística implica un uso del territorio en el tiempo, una convergencia espacio-temporal, una organización y sincronización de flujos a través de estrategias sobre los nodos y las redes. Esta aproximación resalta las limitaciones de miradas parciales y destaca la necesidad de considerar una perspectiva integral que incluya, además, los aspectos organizacionales y de coordinación. La consideración de infraestructuras de circulación; servicios de transporte; infraestructuras de comunicaciones; servicios de almacenamiento y agregado de valor; normativas y regulaciones y costos operativos, entre otros, da cuenta del desafío que involucra una coordinación amplia de políticas estatales de diferente tipo y a cargo de estructuras administrativas diversas. El despliegue de la logística como uno de los “usos del territorio”, abre la posibilidad de acentuar el enfoque territorial. En esta línea de investigación se aborda la optimización de la

logística de distribución de cargas y paquetería. Para tal fin se propone el desarrollo de software logístico, que incorpore herramientas basadas en inteligencia artificial, como soporte para la toma de decisiones a nivel gerencial y asistido por herramientas que permitan evaluar la incidencia de la matriz de costos y buscar el equilibrio del sistema. Se considera que el abordaje de la optimización de la logística de distribución centrada en la provincia de La Pampa tendría impacto en la matriz productiva y de servicios de toda la región.

Palabras claves: Optimización, Logística de distribución, Inteligencia Artificial

Contexto

Esta línea de investigación surge como una proyección del Proyecto "Big data optimization con algoritmos metaheurísticos utilizando frameworks de computación distribuida" que se desarrolla desde el 2021 en la Facultad de Ingeniería de la UNLPam. Este proyecto se lleva adelante en el Laboratorio de Investigación de Sistemas Inteligentes (LISI) de la Facultad de Ingeniería y es dirigido por la Dra. Salto.

En el LISI, desde su creación en 1998, nos hemos abocado al estudio de algoritmos cada vez más eficientes para la solución de problemas complejos, tanto de optimización como de diseño. En este dominio, el objetivo consiste en obtener algoritmos nuevos que den solución al problema y que necesiten un esfuerzo computacional más pequeño que los algoritmos existentes, así como caracterizar su comportamiento para las clases de problemas

que demanda la comunidad científica e industrial en general. Cabe destacar que desde hace varios años, los integrantes de este laboratorio mantienen una importante vinculación con investigadores de las universidades argentinas, Universidad Nacional de San Luis y Universidad Nacional de Río Cuarto, y de la Universidad de Málaga (España), con quienes se realizan trabajos conjuntos.

Introducción

La optimización de la logística de distribución de cargas y paquetería, conocida en el ámbito de la investigación operativa [1] como el problema del enrutamiento de vehículos (VRP, sus siglas en Inglés) [2], es posiblemente uno de los problemas de optimización combinatoria más clásicos que surgen en la cadena logística. El VRP consiste en determinar las rutas que debe tomar una determinada flota de vehículos para recolectar y/o distribuir artículos en ubicaciones conocidas de clientes. Cada artículo suele tener asociado un determinado tamaño o peso. La cantidad total (en términos de peso o tamaño) de las cantidades recolectadas por un solo vehículo no puede exceder su capacidad. En la versión más clásica del VRP, se supone que todos los datos (demandas del cliente, tiempos de viaje, ventanas de tiempo, etc.) se conocen de antemano. El responsable de la toma de decisiones, tanto en grandes compañías como en pequeñas empresas, debe planificar con anticipación las rutas de los vehículos para satisfacer las demandas de los clientes a un costo de viaje mínimo. El costo del viaje puede incluir, entre otros, el tiempo de viaje, las horas hombre, los viáticos, el combustible y los peajes [3], [4]. Desafortunadamente, el VRP es fuertemente NP-duro incluso para un solo objetivo, ya que el problema del viajante de comercio (TSP) [5] puede reducirse polinomialmente a él [6]. Los avances en la tecnología de la comunicación y de la información permiten que las flotas de vehículos sean manejadas en tiempo real. Esto implica usar conjuntamente sistemas de información geográfica, sistemas de posicionamiento global, sensores de flujo de tráfico y teléfonos celulares para obtener datos en tiempo real. Por ejemplo, la ubicación actual de los vehículos, requerimientos de los nuevos

consumidores, el tiempo de viaje en ruta, entre otros. En este contexto, al introducir diversas características dinámicas el VRP clásico se convierte en dinámico (DVRP, por sus siglas en Inglés) [7], [8], representando una parte importante de muchos sistemas de distribución y transporte actuales. El dinamismo puede darse en el bloqueo de la ruta entre dos clientes, en órdenes modificadas por los consumidores, incremento en el tiempo de viaje debido a las malas condiciones climáticas, etc. A estas dos versiones de VRP se suman los requerimientos de nuevos servicios como una mayor atención pública sostenible [9], las restricciones de acceso urbano [10], [11], el aumento de los volúmenes de entrega [12], el uso de instalaciones de transbordo que surgen en la logística de la ciudad [13], la integración de producción, inventario, distribución y enrutamiento de múltiples viajes con ventanas de tiempo [14], la utilización de una flota heterogénea con backhaul (viaje de vuelta) [15], entre otros. Al ser todos estos servicios variantes de VRP pertenecen a la clase de problemas NP-duro. El VRP ha sido objeto de intensos esfuerzos de investigación en el área de la Inteligencia Artificial, considerando enfoques de optimización tanto exactos como heurísticos. Los primeros estudios de solución para el VRP, centrados en técnicas exactas, se pueden encontrar en [16], [17], [18], [19], entre otros. Debido al alto nivel de complejidad de este problema y su amplia aplicabilidad a situaciones de la vida real, las técnicas de solución capaces de producir soluciones de alta calidad en un tiempo limitado, como las heurísticas y las metaheurísticas [20], son de primordial importancia, como lo muestran estas contribuciones: [13], [21], [22], [23], [24], entre muchas otras. Estas técnicas, que pertenecen a la clase de algoritmos de búsqueda estocásticos, reducen de forma significativa la complejidad temporal del proceso de búsqueda aún cuando se ejecutan secuencialmente. Pero la paralelización de este proceso, además de disminuir los tiempos, permite aumentar la calidad de las soluciones halladas. En este sentido, resulta necesario el análisis de las distintas plataformas de hardware (procesadores multicore, unidades de procesamiento gráficas, ambientes cloud) y de software (Hadoop, Spark, MPI, High Processing Computing -HPC- en general) actuales que per-

miten desarrollar técnicas de computación paralelas que mejoran la resolución de problemas con menores tiempos de cómputo. Además de, medir objetivamente el algoritmo en términos de calidad de la solución y tiempo computacional sobre los problemas de los que no se conoce una solución óptima. El desarrollo de estas técnicas, por ser estocásticas, presenta un tercer desafío igual de importante: el uso de métodos y de técnicas estadísticas para valorar la robustez del algoritmo en un amplio rango de casos. Esto conlleva a una mayor usabilidad de esta clase de algoritmos al incorporarlos como módulos de optimización dentro de los sistemas de soporte de decisiones gerenciales [25]. Desde el LISI hemos abordado algunas variantes de este problema [26] así como también del conocido problema del viajante de comercio [27]. Además, hemos resuelto una gran variedad de problemas combinatorios NP-duros con herramientas inteligentes, a saber: diseño de redes de distribución de agua [28], [29], diseño de redes de sensores [30], [31], [32], de planificación de trabajos [33], [34], de ensamblado de fragmentos de ADN [35], [36], corte y empaquetado [37], entre otros. Así como también, la paralelización de estas herramientas en ambientes HPC [38], [29]. La experiencia que hemos alcanzado en la resolución de esta variedad de problemas, nos permitiría aplicar y adaptar estas técnicas en la resolución del problema de optimización de la logística de distribución de cargas y paquetería en la Provincia de La Pampa y su región aledaña. Se trabajará con datos de prueba brindados por la comunidad científica (benchmarks) que nos permitirá contrastar y evaluar nuestra propuesta con algoritmos de la literatura, con el objetivo de medir el desempeño de los algoritmos desarrollados. A partir de esto, podremos utilizar nuestra propuesta algorítmica para optimizar escenarios de distribución reales, incorporando el análisis del impacto de los costos fijos (aquellos que son constantes durante el volumen “normal” de operación del transportista) y variables (aquellos que varían con los servicios o el volumen) propios de nuestra provincia y región aledaña.

Línea de investigación y desarrollo

Para abordar este puntual problema de opti-

mización de la logística de distribución de cargas y paquetería, se propuso el desarrollo de software logístico, que incorpore herramientas basadas en inteligencia artificial, como soporte para la toma de decisiones a nivel gerencial y asistido por herramientas que permitan evaluar la incidencia de la matriz de costos y buscar el equilibrio del sistema. Luego de la identificación y enunciado del VRP y sus variantes, se procede con la formulación de las hipótesis que guiará nuestro trabajo de investigación: H1- La herramienta de software, basada en las técnicas estocásticas de la inteligencia artificial, es capaz de diseñar rutas de distribución eficientes para incrementar la eficiencia operacional, reducir costos y aumentar la satisfacción de los clientes. H2- La determinación de los factores de costos relacionados a cada ruta planificada, su clasificación en cuanto a variabilidad y la asignación de un valor, permiten estudiar la incidencia de cada uno de ellos en el costo total de dicha ruta y contar con información importante para la toma de decisiones. Con el fin de demostrar las hipótesis, el objetivo general de este proyecto es brindar herramientas de software sólidas para la toma de decisiones, que permitan optimizar la logística de distribución de carga y paquetería que minimicen costos e incrementen la eficiencia operacional, no solo en relación al tiempo, sino también en cuanto a los costos en que se incurrirán necesariamente. De esta manera se proporcionan soluciones de alta calidad al problema VRP y sus variantes presentes en nuestra región. Este problema tiene varias aristas a considerar tales como capacidad de carga de los vehículos, una flota heterogénea, rutas abiertas, zonas geográficas de distribución, prioridades de entrega, ventanas de tiempo para contactar a clientes, backhaul, entre otras. También se contempla que dichas situaciones se puedan producir o variar dinámicamente. En función de la alta complejidad de los VRPs es que no se pueden encontrar soluciones exactas en tiempos razonables. Dado lo anterior, los problemas de este tipo se resuelven utilizando técnicas provenientes de la inteligencia artificial como lo son las meta-heurísticas, que buscan encontrar la mejor solución posible en un tiempo acotado. Por lo tanto, el software logístico a desarrollar en este proyecto para la toma de decisiones seguirá tales lineamientos. Para

alcanzar el objetivo general propuesto se establecen los siguientes objetivos específicos: 1- Relevar las características que presenta el VRP en entornos reales de nuestra región. 2- Modelar el problema y formularlo matemáticamente, identificando variables, objetivos y restricciones. 3- Seleccionar el método metaheurístico más adecuado en base al cual se desarrollará el sistema de logística de distribución inteligente. 4- Diseñar e implementar el software logístico que optimizará la logística de distribución. 5- Estudiar el costo de cada ruta de distribución obtenida por el software para verificar su viabilidad en situaciones reales y facilitar la toma de decisiones en forma proactiva. 6- Analizar posibles extensiones distribuidas y/o paralelas de esta herramienta para reducir su tiempo de respuesta.

Resultados esperados

El resultado final de este proyecto es diseñar e implementar un software logístico como soporte para la toma de decisiones a nivel gerencial de empresas que incluyan en sus actividades tareas de distribución de cargas y paquetería. Dicha tecnología permitirá, a partir de la consideración de las distintas variables y restricciones, conseguir el mayor rendimiento y eficacia en los procesos logísticos. Por ello el impacto potencial de esta propuesta cubrirá distintas aristas. Por un lado permitirá la formación específica de un grupo de profesionales y de estudiantes avanzados de las carreras en el abordaje de este complejo problema tan importante para la región. Por otro lado, se desarrollarán aplicaciones apropiadas para potenciales centros de logística y distribución a instalarse en la provincia. La Facultad de Ingeniería se beneficiará, pues este aprendizaje favorece la formación de los recursos humanos en una actividad de transferencia de los resultados de la investigación desarrollada para una futura aplicación del software logístico en una empresa de la región. También es importante resaltar la interdisciplinariedad en el tratamiento de la propuesta ya que permite vincular áreas ingenieriles: Sistemas Inteligentes, Sistemas Distribuidos, Bases de Datos, Ingeniería de Software, Costos Industriales e Investigación Operativa, que

proviene de la Ingeniería en Sistemas y de la Ingeniería Industrial.

Formación de recursos humanos

La realización de este proyecto brindaría el marco adecuado para continuar con la formación de los integrantes del mismo y propiciar un ambiente para la cooperación entre ellos y empresas de la región interesadas en el abordaje de la problemática abordada. Se contempla la incorporación de dos alumnos avanzados de las carreras de Ing. en Sistemas y de Ing. Industrial para que se inicien en la tarea de investigación y sea esta un área propicia para desarrollar sus trabajos de tesis, o bien recientes graduados de las mencionadas carreras que se deseen incorporar a la actividad académica o de investigación. Las tareas a asignar consistirán en: 1 - el desarrollo de una aplicación para interactuar con el software logístico, 2 - la implementación de un entorno gráfico para visualizar las rutas obtenidas en la optimización del problema, 3 - el relevamiento de datos actualizados correspondientes a los costos fijos y variables determinados, y 4 - la verificación de las rutas de distribución obtenidas por la metaheurística con los datos relevados.

REFERENCES

- [1] F. S. Hillier and G. J. Lieberman, *Introducción a la investigación de operaciones*, 9th ed. McGRAW-HILL/INTERAMERICANA EDITORES, 2012.
- [2] G. B. Dantzig and J. H. Ramser, "The truck dispatching problem," *Management Science*, vol. 6, no. 1, pp. 80–91, 1959. [Online]. Available: <https://doi.org/10.1287/mnsc.6.1.80>
- [3] S. Chopra and P. Meindl, *Administración de la Cadena de Suministro: Estrategia Planeación y Operación*, 5th ed. Pearson, 2013.
- [4] R. Ballou, *LOGÍSTICA. Administración de la cadena de suministros*, 5th ed. Pearson, 2004.
- [5] G. Dantzig, R. Fulkerson, and S. Johnson, "Solution of a large-scale traveling-salesman problem," *Journal of the Operations Research Society of America*, vol. 2, no. 4, pp. 393–410, 1954. [Online]. Available: <http://www.jstor.org/stable/166695>
- [6] J. K. Lenstra and A. H. G. R. Kan, "Complexity of vehicle routing and scheduling problems," *Networks*, vol. 11, no. 2, pp. 221–227, 1981. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/net.3230110211>
- [7] H. Psaraftis, *Dynamic Vehicle Routing Problems*. North-Holland, 1988, pp. 223–248.
- [8] O. Madsen, A. Larsen, and M. Solomon, *Recent developments in dynamic vehicle routing systems*, 1st ed. Springer, 2008, vol. 43, pp. 199–220.

- [9] M. Savelsbergh and T. Van Woensel, "50th anniversary invited article—city logistics: Challenges and opportunities," *Transportation Science*, vol. 50, no. 2, pp. 579–590, 2016. [Online]. Available: <https://doi.org/10.1287/trsc.2016.0675>
- [10] S. Ville, J. Gonzalez-Feliu, and L. Dablanc, "The limits of public policy intervention in urban logistics: Lessons from vicenza (italy)," *European Planning Studies*, vol. 21, no. 10, pp. 1528–1541, 2013. [Online]. Available: <https://doi.org/10.1080/09654313.2012.722954>
- [11] R. Elbert and C. Friedrich, "Simulation-based evaluation of urban consolidation centers considering urban access regulations," in *2018 Winter Simulation Conference (WSC)*, 2018, pp. 2827–2838.
- [12] Comisión-Europea, "Directorate general for mobility and transport," 2019. [Online]. Available: <https://www.amt-autoridade.pt/media/1934/2019-transport-in-the-eu-current-trends-and-issues.pdf>
- [13] C. Friedrich and R. Elbert, "Adaptive large neighborhood search for vehicle routing problems with transshipment facilities arising in city logistics," *Computers Operations Research*, vol. 137, p. 105491, 2022.
- [14] B. Ramos, C. Alves, and J. Valério de Carvalho, "An arc flow formulation to the multitrip production, inventory, distribution, and routing problem with time windows," *International Transactions in Operational Research*, vol. 29, no. 1, pp. 526–553, 2022. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/itor.12765>
- [15] W. Wu, Y. Tian, and T. Jin, "A label based ant colony algorithm for heterogeneous vehicle routing with mixed backhaul," *Applied Soft Computing*, vol. 47, pp. 224–234, 2016.
- [16] B. L. Golden and A. A. Assad, "Perspectives on vehicle routing: Exciting new developments," *Operations Research*, vol. 34, no. 5, pp. 803–810, 1986. [Online]. Available: <http://www.jstor.org/stable/170738>
- [17] —, *Vehicle Routing: Methods and Studies*, Amsterdam: North - Holland, 1988.
- [18] M. Desrochers, J. Lenstra, M. Savelsbergh, and F. Soumis, "Vehicle routing with time windows: Optimization and approximation," *Vehicle Routing: Methods and Studies*, vol. 16, 01 1988.
- [19] M. M. SOLOMON and J. DESROSIERS, "Time window constrained routing and scheduling problems," *Transportation Science*, vol. 22, no. 1, pp. 1–13, 1988. [Online]. Available: <http://www.jstor.org/stable/25768291>
- [20] E. Talbi, *Metaheuristics: from Design to Implementation*. Wiley, 2009.
- [21] E. Demir, T. Bektaş, and G. Laporte, "An adaptive large neighborhood search heuristic for the pollution-routing problem," *European Journal of Operational Research*, vol. 223, no. 2, pp. 346–359, 2012.
- [22] G. Dueck, "New optimization heuristics: The great deluge algorithm and the record-to-record travel," *Journal of Computational Physics*, vol. 104, no. 1, pp. 86–92, 1993.
- [23] G. Dueck and T. Scheuer, "Threshold accepting: A general purpose optimization algorithm appearing superior to simulated annealing," *Journal of Computational Physics*, vol. 90, no. 1, pp. 161–175, 1990.
- [24] D. Dumez, F. Lehuédé, and O. Péton, "A large neighborhood search approach to the vehicle routing problem with delivery options," *Transportation Research Part B: Methodological*, vol. 144, pp. 103–132, 2021.
- [25] E. Alba, C. Blum, P. Asasi, C. Leon, and J. Gomez, Eds., *Optimization Techniques for Solving Complex Problems*. Wiley, 2009.
- [26] C. Bermudez, P. Graglia, N. Stark, C. Salto, and H. Alfonso, "Comparison of recombination operators in panmictic and cellular gas to solve a vehicle routing problem," *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, 2010.
- [27] G. F. Minetti, G. Luque, and E. Alba, "The problem aware local search algorithm: an efficient technique for permutation-based problems," *Soft Comput.*, vol. 21, no. 18, pp. 5193–5206, 2017. [Online]. Available: <https://doi.org/10.1007/s00500-017-2515-9>
- [28] C. Bermúdez, C. Salto, and G. Minetti, *Communications in Computer and Information Science*, 2019, ch. Solving the Multi-Period Water Distribution Network Design Problem with a Hybrid Simulated Annealing, pp. 3–16.
- [29] C. Bermudez, H. Alfonso, G. Minetti, and C. Salto, "A parallel optimization solver for the multi-period wdnd problem," *Springer, Cham*, vol. 12886, Sep. 2021.
- [30] J. Hernandez, C. Salto, G. Minetti, M. Carnero, and M. C. Sanchez, "Hybrid simulated annealing for optimal cost instrumentation in chemical plants," *Chemical Engineering Transactions*, vol. 74, pp. 709–714, May 2019.
- [31] G. Minetti, J. Hernandez, M. Carnero, C. Salto, C. Bermúdez, and M. Sánchez, "Tuning a hybrid sa based algorithm applied to optimal sensor network design," *Journal of Computer Science and Technology*, vol. 20, p. e03, 05 2020.
- [32] J. Hernandez, C. Salto, G. Minetti, M. Carnero, and M. Sánchez, "Reliability estimation for sensor networks in chemical plants using monte carlo methods," *Computer Aided Chemical Engineering*, vol. 48, pp. 439–444, 01 2020.
- [33] C. Salto, F. Morero, and C. Bermúdez, "Parallelism and hybridization in differential evolution to solve the flexible job shop scheduling problem," *Journal of Computer Science and Technology*, vol. 20, p. e04, 05 2020.
- [34] G. Minetti, G. Luque, and E. Alba, "The problem aware local search algorithm: an efficient technique for permutation-based problems," *Soft Computing*, vol. 21, 09 2017.
- [35] G. Minetti, E. Alba, and G. Luque, "Seeding strategies and recombination operators for solving the dna fragment assembly problem," *Inf. Process. Lett.*, vol. 108, no. 3, pp. 94–100, 2008.
- [36] G. Minetti and E. Alba, "Metaheuristic assemblers of dna strands: Noiseless and noisy cases," 07 2010, pp. 1–8.
- [37] C. Salto, M. G. Leguizamón, and E. Alba, "Parallel aco algorithms for 2d strip packing," 2010.
- [38] C. Salto, G. Minetti, E. Alba, and G. Luque, *Developing Genetic Algorithms Using Different MapReduce Frameworks: MPI vs. Hadoop: 18th Conference of the Spanish Association for Artificial Intelligence, CAEPIA 2018, Granada, Spain, October 23–26, 2018, Proceedings*, 01 2018, pp. 262–272.

Metaheurísticas paralelas aplicadas al diseño de la red de sensores en plantas químicas

Carlos Bermudez¹, José Luis Hernandez², Mercedes Carnero²,
Gabriela Minetti¹, Carolina Salto¹

¹Laboratorio de Investigación en Sistemas Inteligentes (LISI)
Facultad de Ingeniería - Universidad Nacional de La Pampa
e-mail: ¹{bermudezc, minettig, saltoc, @ing.unlpam.edu.ar}

²Grupo de Optimización (GOp)
Facultad de Ingeniería - Universidad Nacional Río Cuarto
e-mail: ²{jlh, mcarnero@ing.unrc.edu.ar}

Resumen En una planta química inteligente, conocer el estado del proceso en tiempo real es una tarea esencial y que impacta en aspectos tales como la economía, la seguridad y el control. La información es recolectada por sensores distribuidos por toda la planta, encargados de medir y transmitir los valores de variables asociados a temperatura, humedad, presión, composición entre otros. El conjunto de dispositivos utilizados en la medición se denomina red de sensores y su costo de adquisición es una componente importante del costo de capital, lo cual justifica la importancia de formular el diseño de la red de sensores como un problema de optimización y de contar con herramientas sistemáticas para su resolución. Por este motivo, esta línea de investigación se enfoca en el diseño y desarrollo de un algoritmo de optimización inteligente basado en un *Simulated Annealing* híbrido y paralelo (PHSA, por sus siglas en Inglés), aplicado a la resolución de una red de instrumentación de costo mínimo. El paralelismo se aplica a nivel algorítmico, siguiendo un modelo cooperativo. Entre los parámetros de migración, los criterios de reemplazo tienen un papel importante en el desempeño de PHSA. Nuestro principal objetivo es analizar el comportamiento de PHSA considerando diferentes criterios de reemplazo. Los resultados obtenidos por PHSA alcanza la mejor solución conocida para casos de redes de sensores grandes y complejas.

Palabras claves: Paralelismo, Metaheurísticas, Optimización, Red de sensores, Simulated Annealing, Hibridación

CONTEXTO

Esta línea de investigación se desarrolla en el marco de dos proyectos de investigación, acreditados en distintas universidades argentinas. Uno llevado a cabo en el Grupo de Optimización (GOp), de la Facultad de Ingeniería de la Universidad Nacional de Río Cuarto. El segundo se desarrolla en el Laboratorio de Investigación de Sistemas Inteligentes (LISI), de la Facultad de Ingeniería de la Universidad Nacional de La Pampa.

El diseño de la red de sensores (SNDP, por sus siglas en Inglés) es considerado como elemento clave para hacer realidad la planta inteligente ([1]), ya que proporciona un conocimiento integral del estado actual del proceso. De manera general, el diseño sistemático de una red de sensores en una planta se formula como un problema de optimización ([2], [3]). Consiste en determinar para cada variable si debe medirse o no al tiempo que se optimiza un criterio de desempeño particular. Estos últimos pueden clasificarse como criterios asociados con la performance del sistema de sensores propiamente dicho ([4], [5]) o relacionados con indicadores del proceso para el cual está diseñada la instrumentación ([6], [7], [8]).

Para un problema dado, esto es una función objetivo y un conjunto de restricciones específicas, la cantidad de posibilidades en relación a cuántas y cuáles variables medir puede alcanzar valores muy altos incluso para plantas pequeñas, con un costo computacional asociado que puede resultar prohibitivo. Por ende, un aspecto importante se vincula

con la elección de la metodología de resolución, que se puede clasificar en dos grandes grupos: métodos exactos y métodos heurísticos. El análisis del rendimiento exhibido por los algoritmos exactos ([9], [3]) muestra que son dependientes del problema y de su tamaño. Es decir, no son lo suficientemente robustos como para resolver un amplio rango de especificaciones, ni tienen posibilidad de escalamiento. La desventaja más notoria de los algoritmos exactos es que en muchos casos no pueden garantizar la optimalidad. Esta limitación proviene principalmente de la naturaleza combinatoria del problema, lo que resulta en la imposibilidad de contar con algoritmos de tiempo polinomial para los casos de interés práctico. Por lo tanto, los métodos de optimización heurística emergen como la alternativa de solución más factible para abordar diseños de mayor dimensión ([10], [11], [12]).

El proceso de optimización en el diseño de SN para instancias de gran tamaño puede requerir de estrategias para reducir el tiempo de ejecución del proceso. En consecuencia, se utilizan técnicas de paralelización sobre los métodos heurísticos ([13]), consiguiendo una mejora sustancial en el tiempo dedicado a dicha optimización.

En esta línea de investigación se propone un algoritmo *Simulated Annealing* (SA) ([14], [15]) paralelo para optimizar el SNDP. En este modelo, muchos algoritmos *Simulated Annealing* híbridos (HSA) ([16]) se lanzan en paralelo e intercambian información para mejorar la calidad de las soluciones y aumentar la eficiencia. En consecuencia, la principal contribución de nuestra investigación es el desarrollo de un algoritmo inteligente basado en un SA híbrido y paralelo, denominado PHSA, que se utilice para la toma de decisiones durante el diseño de redes de sensores complejas en grandes plantas químicas.

1. DESARROLLO

En esta sección se describe la línea de investigación ya mencionada, introduciendo una explicación detallada del problema y del algoritmo propuesto para resolverlo.

A. Problema de diseño de la red de sensores

El SNDP se puede definir como el problema de encontrar una estructura de sensores que satisfaga

un conjunto de restricciones a la vez que se optimiza un criterio de desempeño determinado. Una formulación particular en la que se ha trabajado es aquella donde se pretende diseñar una red de instrumentación de costo mínimo que satisfaga restricciones de precisión y estimabilidad sobre un conjunto de variables claves. Formalmente, el problema de SNDP se puede expresar como se muestra en la Ecuación (1), donde \mathbf{q} es un vector n -dimensional de variables binarias tal que $q_i = 1$ si se mide la variable i y $q_i = 0$ en caso contrario; \mathbf{c}^T es el vector de costo; $\hat{\sigma}_k$ es el desvío estándar de la variable k contenida in S_σ después de aplicar un procedimiento de conciliación de datos ([17]), y E_l representa el grado de estimabilidad de la variable l -ésima incluida en S_E . Además, S_σ y S_E son el conjunto de variables clave del proceso con requisitos de precisión y capacidad para estimarse, respectivamente.

$$\min \mathbf{c}^T \mathbf{q} \quad (1)$$

sujeto a:

$$\begin{aligned} \hat{\sigma}_k(\mathbf{q}) &\leq \sigma_k^*(\mathbf{q}) & \forall k \in S_\sigma \\ E_l &\geq 1 & \forall l \in S_E \\ \mathbf{q} &\in \{0, 1\}^n \end{aligned}$$

En esta formulación, se supone que un modelo algebraico lineal representa la operación de la planta, las mediciones están sujetas a errores aleatorios no correlacionados, solo hay un dispositivo de medición potencial para cada variable y no hay restricciones para la localización de instrumentos. Con respecto al grado de restricción de estimabilidad, si $E_l=1$, la factibilidad de la restricción se puede comprobar mediante la ejecución de un procedimiento de clasificación de variables, que se puede lograr mediante la proyección de matrices, la descomposición QR o la cooptación de matrices ([18], [19]).

B. Algoritmo PHSA

PHSA se basa en HSA, (*Hybrid Simulated Annealing*) ([16]), que funciona como metaheurística principal con una búsqueda local ad hoc subordinada, inspirada en la búsqueda tabú con la técnica de oscilación estratégica, SOTS. El algoritmo PHSA

propuesto sigue un modelo de paralelización a nivel algorítmico (ver [13]). PHSA consiste en n HSAs ejecutados en paralelo, que intercambian información para mejorar la calidad de las soluciones y la eficiencia. Por lo tanto, el diseño de PHSA sigue una estrategia de búsqueda cooperativa.

Cada HSA genera su solución inicial S_0 y se ejecuta de forma independiente, intercambiando información entre ellos para encontrar soluciones mejores y más robustas. Se pueden usar diferentes topologías de interconexión, pero nos enfocamos en la topología de anillo, por lo que llamamos a nuestra propuesta PHSA_ring; donde los n HSAs se organizan en un anillo lógico unidireccional. El i -ésimo HSA envía su solución actual S_1 a su vecino, el $(i+1)$ -ésimo HSA, en el anillo con una frecuencia determinada y después de un intercambio asíncrono. El HSA que recibe una solución (S_2) utiliza un determinado criterio de reemplazo para aceptarla o no; luego continúa con su proceso de búsqueda independiente.

En esta investigación, consideramos tres diferentes criterios de reemplazo ([20]), denominados *RC1*, *RC2* y *RC3*.

- *RC1*. S_2 se acepta con la probabilidad de Boltzmann.
- *RC2*. S_2 se acepta si es mejor que la solución actual (S_1).
- *RC3*. S_2 se acepta si es mejor que la solución local mejor encontrada (S_b).

2. RESULTADOS OBTENIDOS

El desempeño de PHSA_ring se evalúa considerando procesos químicos de alta complejidad y tamaño. Como primer caso de estudio, se abordó una planta de etileno simplificada, compuesta por 47 unidades y 82 corrientes, cuyo funcionamiento solo se representa a través de balances de masa globales. En segundo lugar, se utilizó el Proceso Eastman de Tennessee (TEP) ([21]), un problema ampliamente considerado en la literatura de monitoreo y control de procesos químicos. Tanto los balances de masa globales como los de componentes se consideran en el TEP. El sistema completo comprende 42 ecuaciones linealizadas alrededor del punto de operación y un total de 78 variables. En el primer caso de estudio, se observa un conjunto de 14 variables requeridas con restricciones de precisión en 6 de ellas. El

segundo caso de estudio considera un conjunto de 24 variables requeridas, todas ellas con restricciones de precisión. La desviación estándar de los caudalímetros es del 2% de los caudales reales correspondientes para ambos casos.

El desempeño de PHSA_ring se evalúa, empíricamente y con soporte estadístico, en términos de calidad de la solución, tiempo de ejecución y cantidad de evaluaciones ([22]). En función a lo evaluado, se observó que PHSA_ring encontró los mejores costos de solución conocidos. Como consecuencia, un modelo basado en HSA cooperativos a nivel algorítmico resulta ser una excelente opción para resolver este problema. Las diferentes pruebas estadísticas demostraron que el *RC3* permitió mejorar significativamente la calidad de la solución. Finalmente, al evaluar la eficiencia de PHSA_ring, al considerar las tasas de aciertos y las medidas de *speedup* [23], se confirma que PHSA_ring mantuvo la propiedad de escalabilidad porque su rendimiento mejoró al aumentar la cantidad de HSA y los valores de *speedup* fueron casi lineales.

3. FORMACIÓN DE RECURSOS HUMANOS

Cada año se incorporan a los proyectos alumnos avanzados en la carrera Ingeniería en Sistemas, quienes trabajan en temas relacionados a la resolución de problemas de optimización usando técnicas inteligentes, con el objeto de guiarlos en el desarrollo de sus tesis de grado y, también, de formar futuros investigadores científicos. Por otra parte, los docentes-investigadores que integran los proyectos realizan diversos cursos de posgrado relacionados con la temática del proyecto, con el objetivo de sumar los créditos necesarios para cursar carreras de posgrado.

REFERENCES

- [1] P. Christofides, J. Davis, N. El-Farra, D. Clark, H. K., and J. Gipson, "Smart plant operations: Vision, progress and challenges." *AICHE J*, vol. 53, pp. 2734–2741, 2007.
- [2] M. Bagajewicz, "Design and retrofit of sensor networks in process plants," *AICHE Journal*, vol. 43, no. 9, pp. 2300–2306, 1997.
- [3] D. Nguyen and M. Bagajewicz, "New efficient breadth-first/level traversal tree search method for the design and upgrade of sensor networks," *AICHE Journal*, vol. 57, no. 5, pp. 1302–1309, 2011.

- [4] M. Carnero, J. L. Hernández, and M. Sánchez, "Optimal sensor location in chemical plants using the estimation of distribution algorithms," *Industrial & Engineering Chemistry Research*, vol. 57, no. 36, pp. 12 149–12 164, 2018.
- [5] P. Kotecha, M. Bhushan, R. Gudi, and M. Keshari, "A duality based framework for integrating reliability and precision for sensor network design," *Journal of Process Control*, vol. 18, no. 2, pp. 189–201, 2008.
- [6] P. Sen, K. Sen, and U. M. Diwekar, "A multi-objective optimization approach to optimal sensor location problem in IGCC power plants," *Applied Energy*, vol. 181, pp. 527–539, 2016.
- [7] P. Paul, D. Bhattacharyya, R. Turton, and S. E. Zitney, "Sensor network design for maximizing process efficiency: an algorithm and its application," *AIChE Journal*, vol. 61, no. 2, pp. 464–476, 2015.
- [8] G. Sambito, M. and Freni, "Strategies for improving optimal positioning of quality sensors in urban drainage systems for non-conservative contaminants," *Water*, vol. 13, pp. 1–14, 2021.
- [9] J. Zhang and D. J. Chmielewski, "Profit-based sensor network design using the generalized benders decomposition," in *2017 American Control Conference (ACC)*, 2017, pp. 3894–3899.
- [10] J. Hernandez, C. Salto, G. Minetti, M. Carnero, and M. C. Sanchez, "Hybrid simulated annealing for optimal cost instrumentation in chemical plants," *Chemical Engineering Transactions*, vol. 74, pp. 709–714, May 2019.
- [11] Y.-J. He and Z.-F. Ma, "Optimal design of linear sensor networks for process plants: A multi-objective ant colony optimization approach," *Chemometrics and Intelligent Laboratory Systems*, vol. 135, pp. 37–47, 2014.
- [12] Á. Panizo, G. Bello-Organ, M. Carnero, J. Hernández, M. Sánchez, and D. Camacho, "An artificial bee colony algorithm for optimizing the design of sensor networks," in *Intelligent Data Engineering and Automated Learning – IDEAL 2018*. Cham: Springer International Publishing, 2018, pp. 316–324.
- [13] E. Talbi, *Metaheuristics: From Design to Implementation*. Wiley Publishing, 2009.
- [14] S. Kirkpatrick, C. G. Jr, and M. Vecchi, "Optimization by simulated annealing," *Science*, no. 220, pp. 671–680, 1983.
- [15] V. A. Cicirello, "Variable annealing length and parallelism in simulated annealing," *CoRR*, vol. abs/1709.02877, 2017.
- [16] J. L. Hernández, C. Salto, G. F. Minetti, M. Carnero, C. Bermúdez, and M. Sánchez, "Tuning a hybrid sa based algorithm applied to optimal sensor network design," *Journal of Computer Science & Technology*, vol. 20, no. 1, 2020.
- [17] M. Bagajewicz and M. Sanchez, "Reallocation and upgrade of instrumentation in process plants," *Computers & Chemical Engineering*, vol. 24, no. 8, pp. 1945 – 1959, 2000.
- [18] J. Romagnoli and M. Sanchez, *Data Processing and Reconciliation for Chemical Process Operations*. Academic Press: San Diego, CA, 2000.
- [19] S. Narasimhan and C. Jordache, *Data Reconciliation and Gross Error Detection*. Gulf Publishing Company, 2000.
- [20] C. Bermudez, H. Alfonso, G. Minetti, and C. Salto, "A parallel optimization solver for the multi-period wond problem," in *Hybrid Artificial Intelligent Systems*, H. Sanjurjo González, I. Pastor López, P. García Bringas, H. Quintián, and E. Corchado, Eds. Cham: Springer International Publishing, 2021, pp. 625–636.
- [21] J. Downs and E. Vogel, "A plant-wide industrial process control problem," *Comp. Chem. Eng.*, vol. 17, pp. 245–255, 1993.
- [22] C. Bermudez, H. Alfonso, G. Minetti, and C. Salto, "A parallel optimization solver for the multi-period wond problem," in *Hybrid Artificial Intelligent Systems*, H. Sanjurjo González, I. Pastor López, P. García Bringas, H. Quintián, and E. Corchado, Eds. Cham: Springer International Publishing, 2021, pp. 625–636.
- [23] E. Alba, "Parallel evolutionary algorithms can achieve super-linear performance," *Inf. Process. Lett.*, vol. 82, pp. 7–13, 2002.

GESTIÓN SUSTENTABLE DEL TRÁNSITO EN CIUDADES INTELIGENTES Y SOSTENIBLES

Villagra A., Pandolfi D., Mercado V., Ramos L., Torres M., Del Do M., Molina D.,
Varas V.

Laboratorio de Tecnologías Emergentes (LabTEM)
Instituto de Tecnología Aplicada (ITA) - Unidad Académica Caleta Olivia
Universidad Nacional de la Patagonia Austral

{avillagra, dpandolfi, vmercado,lramos}@uaco.unpa.edu.ar, marianagalos@gmail.com,
madeldo@gmail.com, {dmolina,vvaras}@uaco.unpa.edu.ar

RESUMEN

Los sistemas de movilidad urbana son variados y han evolucionado rápidamente con nuevos servicios para adaptarse a una demanda creciente, considerando los cuatro requisitos modernos para ciudades inteligentes: ser inclusivos, seguros, resilientes y sostenibles. Estos nuevos servicios tratan en gran medida con vehículos, semáforos y otros "objetos" similares. Sin embargo, los problemas de tránsito aumentan constantemente y las ciudades del futuro solo pueden ser verdaderamente inteligentes si habilitan la Movilidad Inteligente (en inglés, *Smart Mobility*). Mejorar la movilidad representa un gran reto.

En esta línea de trabajo se presenta una propuesta de investigación enfocada en los desafíos relacionados con la movilidad. Se hace desde la perspectiva de construir nuevos prototipos basados en sistemas inteligentes, mejorados con metodología y tecnologías diferentes con el fin de exhibir "inteligencia holística". En particular se pretende contribuir a la reducción del consumo de combustibles, la emisión de gases, y el tiempo de traslado, así como al mejoramiento de la calidad del aire. Se utiliza la infraestructura existente (semáforos), optimizando el flujo de vehículos sin ningún coste adicional, y sin requerir el uso de aplicaciones especializadas por parte de los conductores.

Palabras clave: Ciudades Inteligentes, Optimización, Planificación semafórica, Movilidad Inteligente.

CONTEXTO

La línea de trabajo se lleva a cabo en el Laboratorio de Tecnologías Emergentes (LabTEM), Instituto de Tecnología Aplicada (ITA) de la Unidad Académica Caleta Olivia Universidad Nacional de la Patagonia Austral, en el marco del Proyecto de Investigación 29/B273 "Ciudades inteligentes y sostenibles: iniciativas y desafíos". Este proyecto se desarrolla en cooperación con el LIDIC de la UNSL, y el Grupo NEO de la UMA (España).

1. INTRODUCCIÓN

Las aplicaciones de movilidad inteligente, como el control de semáforos, estacionamientos inteligentes y la gestión del tráfico rodado, están empezando a implantarse en todo el mundo y aportan beneficios a las ciudades, como una mejor calidad de vida, una reducción de costes, un uso más eficiente de la energía y una disminución de las emisiones de los vehículos. Los semáforos son elementos simples pero esenciales utilizados en entornos urbanos para organizar el tránsito, principalmente de vehículos [1], para evitar accidentes y mejorar el flujo de tráfico. Una ciudad con tráfico congestionado y muchos

atascos conlleva, sin duda, a más contaminación, ya que los vehículos se ven obligados a hacer muchas paradas o a mantener el coche en marcha durante más tiempo y, en consecuencia, emiten grandes cantidades de dióxido de carbono (CO₂). Por lo tanto, con una gestión inteligente del tráfico se pueden reducir drásticamente los tiempos de viaje, el consumo de combustible y las emisiones contaminantes. El problema del control del tráfico por zonas es uno de los más difíciles en el campo del control del flujo de tráfico y requiere la utilización de técnicas avanzadas para resolver incluso los casos más pequeños. A diferencia de técnicas como los semáforos inteligentes, que requieren nuevas infraestructuras, la ubicación de sensores y modificaciones en obras civiles ya estructuradas, el uso de técnicas de inteligencia artificial en la optimización de los ciclos de los semáforos se presenta como una herramienta viable, rápida, eficiente y de bajo coste [2].

En los últimos años, las metaheurísticas se han utilizado ampliamente para abordar este problema, especialmente el algoritmo genético (GA), la optimización por colonia de hormigas (ACO) y la optimización por enjambre de partículas (PSO) [3-14]. Los autores en [2] y [14] han utilizado PSO para construir programas de ciclos de semáforos, con el objetivo de reducir el consumo de combustible y las emisiones de los vehículos en las áreas metropolitanas. En [15] se presenta un enfoque paralelo para el problema de programación de semáforos mediante el uso de la Evolución Diferencial (ED), aplicada a escenarios de problemas cercanos a la realidad que consisten en dos grandes áreas urbanas ubicadas en las ciudades de Málaga, España, y París, Francia. Sus resultados superan a los obtenidos por un algoritmo PSO en términos de calidad de la solución.

Para abordar la movilidad inteligente será necesario la aplicación de técnicas de vanguardia de diferentes dominios que permitan flexibilidad, autoadaptabilidad, robustez, alta dimensionalidad (escalabilidad) y eficiencia.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

En esta sección se describe la línea de investigación que se lleva a cabo en el proyecto.

En ciudades grandes son objetivos comunes controlar, disciplinar el tránsito y reducir accidentes en la ciudad, invirtiendo en sistemas de monitoreo y administración de tráfico. Los resultados que deben lograrse incluyen, por ejemplo, el uso de radares de velocidad, la programación adaptativa y en tiempo real de los semáforos, tomando en cuenta, entre otros factores, la concentración y el flujo de vehículos (dando prioridad a las ambulancias y los vehículos policiales y un corredor preferencial para colectivos), la concentración de peatones y la velocidad de los vehículos.

Otra fuente de preocupación común es la oferta de sistemas de transporte público más eficiente, adecuado al desarrollo urbano y a la equidad social en relación con los desplazamientos. Muchas de las soluciones tienen el objetivo de preparar a la ciudad para la implementación, en el futuro, de un sistema multimodal de transporte, que incluya diferentes medios (bicicleta, subtes, franjas exclusivas para colectivos, vehículos livianos sobre rieles), y contribuya así a la reducción del consumo de combustibles, la emisión de gases y el tiempo de traslado, así como al mejoramiento de la calidad del aire.

La optimización de la movilidad inteligente surgió para reducir la contaminación generada por el tráfico [16], con excelentes resultados (hasta un 50% de reducción). Una manera de reducir el impacto negativo del tráfico consiste en optimizar las rutas. Cada ruta tiene costos asociados: tiempo, dinero, contaminación. En consecuencia, se debe optimizar simultáneamente más de un objetivo (por ejemplo, tiempo versus contaminación) [17-20]. Además, los datos utilizados en la optimización no son precisos (contienen errores) y varían durante un viaje. Existen trabajos que se centran en proporcionar rutas

personalizadas a las necesidades de los ciudadanos y modificarlas según el estado actual de las rutas y el tráfico

Para el desarrollo de esta línea seguiremos el método científico, a fin de analizar cuestiones de investigación abiertas en estos campos, definir los desafíos y validar nuestros resultados con prototipos reales. Dar soluciones significa resolver muchos problemas tecnológicos, necesidad de plataformas de alto rendimiento, y visualización intuitiva (mapas digitales y gemelo). Las tecnologías necesarias son muchas y modernas, pero serán inútiles sin una inteligencia real capaz de extraer información de extensos conjuntos de datos junto con una gran cantidad de complejos problemas de optimización por resolver.

Nuestra hipótesis es que creando nuevos algoritmos bioinspirados podremos cumplir con los requisitos, especialmente cuando se combinan con técnicas de aprendizaje automático para ofrecer una neuro-evolución rápida, algoritmos paralelos multiobjetivo, selección de características y metaheurísticas dinámicas, por nombrar algunos. Además, la incorporación de funciones subrogadas aportará mayor eficiencia computacional al proceso de optimización.

3. RESULTADOS

OBTENIDOS/ESPERADOS

Hemos abordado el problema de programación de semáforos con dos versiones de un Algoritmo Genético Celular, cGA (síncrono y asíncrono) para resolver instancias grandes y reales. Nuestros algoritmos superan las técnicas de vanguardia y las configuraciones de expertos. Se llevaron a cabo varios análisis en profundidad de los resultados, estudio genotípico y fenotípico (para mayor detalle ver [20]).

La investigación actual y futura se centra en lograr disminuir el número de evaluaciones con el

objetivo de reducir el esfuerzo computacional en términos de tiempo de procesamiento.

Planeamos definir una función de aptitud más precisa que explore el espacio de búsqueda de manera más eficiente.

Aplicar diferentes modelos de funciones subrogadas utilizando redes neuronales artificiales (ANN) y otros modelos de SMT (*Surrogate Modeling Toolbox*). Integrar los modelos con el cGA y analizar resultados.

Tenemos previsto modelar diferentes problemas con datos de varias ciudades para nuestros algoritmos de optimización., basándonos en trabajos como [18 y 19] donde se realiza la importación de mapas de *OpenStreetMap* en el simulador de tráfico SUMO. Los mapas realistas incluirán rotondas reales, semáforos, etc. Además, utilizaremos datos abiertos publicados por las autoridades locales, para mejorar la precisión. También se pretende desarrollar varios prototipos y servicios web para poner en práctica nuestras ideas de movilidad inteligente, ofreciendo servicios web para optimizar las rutas.

Finalmente, se procurará colaborar con los gobiernos locales para la implementación de políticas y acciones inteligentes y sostenibles que impacten en la calidad de vida de los ciudadanos.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo se encuentra formado por una doctora y dos Magister en Ciencias de la Computación, un Magister en Matemática Avanzada, cuatro Ingenieros en Sistemas cursando estudios de Posgrado.

Este proyecto de investigación proporcionará un marco propicio para la iniciación y/o finalización de estudios de posgrado de los integrantes docentes.

5. BIBLIOGRAFIA

- [1] Papageorgiou, M., Diakaki, C., Dinopoulou, D., Kotsialos, A., Wang, Y. Review of road traffic control strategies, *Proc. IEEE* 91 (2003) 2043–2067.
- [2] Garcia-Nieto, J., Ferrer, J., Alba, E. Optimising traffic lights with metaheuristics: reduction of car emissions and consumption, in: *International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2014, pp. 48–54.
- [3] Celtek, S. A., Durdu, A., & Ali, M. E. M. (2020). Real-time traffic signal control with swarm optimization methods. *Measurement*, 166, 108206.
- [4] Shaikh, P. W., El-Abd, M., Khanafer, M., & Gao, K. (2020). A review on swarm intelligence and evolutionary algorithms for solving the traffic signal control problem. *IEEE transactions on intelligent transportation systems*, 23(1), 48–63.
- [5] Sánchez, J., Galán, M., Rubio, E., Applying a traffic lights evolutionary optimization technique to a real case: “Las Ramblas” area in Santa Cruz de Tenerife, *IEEE Trans. Evol. Comput.* 12 (2008) 25–40.
- [6] Stolfi, D. H., & Alba, E. (2021). Yellow Swarm: LED panels to advise optimal alternative tours to drivers in the city of Malaga. *Applied Soft Computing*, 109, 107566.
- [7] Baskan, O., Haldenbilen, S., Ant colony optimization approach for optimizing traffic signal timings, in: *Ant Colony Optimization-Methods and Applications*, InTech, 2011.
- [8] Nguyen, T. H., & Jung, J. J. (2021). Ant colony optimization-based traffic routing with intersection negotiation for connected vehicles. *Applied Soft Computing*, 112, 107828.
- [9] Jabbarpour, M., Malakooti, H., Noor, R., Anuar, N., Khamis, N., Ant colony optimisation for vehicle traffic systems: applications and challenges, *Int. J. Bio-Inspir. Comput.* 6 (2014) 32–56.
- [10] Yadav, A., & Nuthong, C. (2020, May). Traffic signal timings optimization based on genetic algorithm and gradient descent. In *2020 5th International Conference on Computer and Communication Systems (ICCCS)* (pp. 670–674). IEEE.
- [11] Cui, C.-Y., Lee, H.-H., Distributed traffic signal control using PSO based on probability model for traffic jam *Intelligent Autonomous Systems*, 12, Springer, 2013, pp. 629–639.
- [12] García-Nieto, J., Alba, E., Olivera, A., Swarm intelligence for traffic lights scheduling: application to real urban areas, *Eng. Appl. Artif. Intell.* 25 (2012) 274–283.
- [13] Garcia-Nieto, J., Olivera, A., Alba, E., Optimal cycle program of traffic lights with particle swarm optimization, *IEEE Trans. Evol. Comput.* 17 (2013) 823–839.
- [14] Olivera, A., García-Nieto, J., Alba, E., Reducing vehicle emissions and fuel consumption in the city by using particle swarm optimization, *Appl. Intell.* 42(2015) 389–405.
- [15] Souravlias, D., Luque, G., Alba, E., Parsopoulos, K., Smart traffic lights: a first parallel computing approach, in: *Intelligent Networking and Collaborative Systems (INCoS)*, International Conference on Intelligent Networking and Collaborative Systems (INCoS), IEEE, 2016, pp. 229–236.
- [16] Shaheen, S., Rodier, C. J., & Finson, R. S. (2003). University of California, Davis Long-Range Development Plan: A Davis Smart Mobility Model. California Partners for Advanced Transit and Highways (PATH).
- [17] Arnott, R., Rave, T., & Schöb, R. (2005). Alleviating urban traffic congestion. MIT Press Books, 1.
- [18] Stolfi, D. H., & Alba, E. (2014). Red Swarm: Reducing travel times in smart cities by using bio-inspired algorithms. *Applied Soft Computing*, 24, 181–195.
- [19] Stolfi, D. H., & Alba, E. (2014, July). Eco-friendly reduction of travel times in european smart cities. In *Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation* (pp. 1207–1214).
- [20] Villagra, A., Alba, E., & Luque, G. (2020). A better understanding on traffic light scheduling: New cellular GAs and new in-depth analysis of solutions. *Journal of Computational Science*, 41, 101085.

DETECCIÓN DE LA CALIDAD DEL AGUA MEDIANTE IMÁGENES SATELITALES: REVISIÓN SISTEMÁTICA DE LA LITERATURA CON ANÁLISIS CUANTITATIVO

M. Silvia Vera Laceiras, Norcelo G. De Miranda, Horacio Kuna, Eduardo Zamudio

Instituto de Investigación, Desarrollo e Innovación en Informática (IIDII)

Facultad de Ciencias Exactas, Químicas y Naturales, Universidad Nacional de Misiones

[vlhsilvia, hdkuna, eduardozamudi}@fceqyn.unam.edu.ar](mailto:{vlhsilvia, hdkuna, eduardozamudi}@fceqyn.unam.edu.ar)

norcelodemiranda@gmail.com

RESUMEN

Este artículo tiene como objetivo presentar avances y la incorporación de aplicación de imágenes satelitales al equipo interdisciplinario de la línea de investigación “CIENCIA DE DATOS COMO HERRAMIENTA DE SOPORTE EN LA GESTIÓN PÚBLICA DE CALIDAD DEL AGUA” con las imágenes satelitales se pretende generar datos crudos sobre el río Paraná, para la ingesta de información y aplicación con ciencia de datos. Contribuyendo a través de este medio al conocimiento del dominio, y aportar a la toma de decisiones. Como tarea inicial, se buscó la familiarización de los términos y variables utilizadas en este contexto a través de un metaanálisis. La RSL cuantitativa: “Detección de calidad de agua mediante imágenes obtenidas a través de teledetección satelital”. Como resultado otorga un número de estudios que ofrecen una perspectiva representativa del conjunto de publicaciones sobre el tema respondiendo a preguntas y subpreguntas de investigación diseñadas por los investigadores. El resultado permite un conocimiento de la terminología básica utilizada, y los aspectos relevantes sobre la calidad del agua y su relación con algas tóxicas, microplásticos y sólidos en suspensión.

CONTEXTO

Esta línea de investigación se incorpora dentro del proyecto de investigación “CIENCIA DE DATOS COMO HERRAMIENTA DE SOPORTE EN LA GESTIÓN PÚBLICA DE CALIDAD DEL AGUA” número 16/Q1224-IDP del Instituto de Investigación, Desarrollo e Innovación en Informática (IIDII) de la Facultad de Ciencias Exactas, Químicas y Naturales (FCEQyN) de la Universidad

Nacional de Misiones (UNaM). Realiza su aporte a partir del estudio de imágenes satelitales del Río Paraná de dominio público o semipúblico y generando datos crudos para su tratamiento con ciencia de datos. Obteniendo datos que sirvan de ingesta en la gestión de conocimiento de la calidad de agua, y ayude a determinar cómo puede verse afectada nuestra zona por algas tóxicas y microplástico.

INTRODUCCIÓN

Los Objetivos de Desarrollo Sostenible (ODS) son un conjunto de acciones de los países para preservar nuestros océanos y bosques, reducir la desigualdad y estimular el crecimiento económico. Los ODS de gestión del agua exigen un seguimiento constante de las métricas de cobertura de la calidad del agua. Estas métricas aseguran “Garantizar la disponibilidad de agua y su gestión sostenible y el saneamiento para todos” [1] a través de acciones concretas como que de aquí a 2030, se propone mejorar la calidad del agua reduciendo la contaminación, eliminando el vertimiento y minimizando la emisión de productos químicos y materiales peligrosos, reduciendo a la mitad el porcentaje de aguas residuales sin tratar y aumentando considerablemente el reciclado y la reutilización sin riesgos a nivel mundial.

El tema de estudio es la calidad del agua a través de la observación y medición de sus contaminantes, usando para este fin el análisis de datos obtenidos a través de teledetección satelital de imágenes.

A fin de entender el alcance de los términos que se usarán en adelante, se buscan los estudios adecuados para realizar un metaanálisis o Revisión sistemática de la Literatura (RSL) cuantitativa de los estudios

obtenidos utilizando metodología estadística especializada. Respondiendo a la pregunta de investigación ¿Para qué aplicamos teledetección en calidad de agua? y a las subpreguntas de investigación ¿Qué indica la presencia de algas tóxicas con respecto a la calidad del agua?, ¿Qué indica la presencia de microplásticos con respecto a la calidad del agua?, ¿Qué índices se pueden generar a través de la teledetección de microplásticos o algas en los cursos de agua?

En Línea de investigación y desarrollo se explica el objetivo del artículo y cuál es el procedimiento implementado para obtener los resultados. En la sección Análisis de resultados se comparten los resultados obtenidos con el procedimiento. Finalmente, en la sección Formación de recursos humanos.

LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Este artículo presenta avances y la incorporación de imágenes satelitales al proyecto de investigación “CIENCIA DE DATOS COMO HERRAMIENTA DE SOPORTE EN LA GESTIÓN PÚBLICA DE CALIDAD DEL AGUA”. Con las imágenes satelitales se agregan a través de transformaciones específicas, datos crudos obtenidos del procesamiento digital de estas imágenes, en tomas puntuales del río Paraná. para la ingesta de información y aplicación en ciencia de datos. Contribuyendo a través de este medio al conocimiento del dominio, y aportar a la toma de decisiones . cómo tarea inicial, se buscó la familiarización de los términos y variables utilizadas en este contexto a través de un metaanálisis.

El metaanálisis de los resultados permite resumir en un solo valor numérico toda la evidencia relacionada con un tema puntual, aumentando la potencia estadística y la precisión del estimador puntual [2] Y se refiere al análisis estadístico de los datos de estudios primarios independientes enfocados en la misma pregunta, que tiene como objetivo generar una estimación cuantitativa del fenómeno estudiado, por ejemplo, la efectividad de la intervención [3]. Dichos

análisis son esencialmente observacionales y utilizan estudios como unidad de investigación. Si bien puede ser controversial, pues si los estudios seleccionados tienen algún sesgo también lo tendrá la conclusión, la fuerza del metaanálisis radica en la capacidad de resumir un gran volumen de literatura en una sola publicación y producir conclusiones relevantes. El metaanálisis permite combinar los resultados de varios informes diferentes en un informe, para crear una estimación única y más precisa de un efecto [4]. Los análisis estadísticos en un metaanálisis están guiados por un modelo estadístico que debe asumirse previamente, para responder a preguntas planteadas con antelación La tarea principal del modelo estadístico es establecer las propiedades de la población del tamaño del efecto a partir de la cual se estima el tamaño del efecto individual [4,5].

Como objetivo del metaanálisis se puede “aumentar el poder estadístico; lidiar con la controversia cuando los estudios individuales no están de acuerdo [6,7]. También posee ventajas como que permite a los investigadores agrupar datos de muchos ensayos demasiado pequeños por sí mismos para permitir conclusiones seguras. Aunque idealmente cualquier ensayo clínico debería planificar un tamaño de muestra adecuado, históricamente la mayoría de los ensayos no han tenido el poder estadístico suficiente.

En 2002, un estudio de 5503 ensayos clínicos [8] identificó que el 69% tenía menos de 100 sujetos. Los ensayos pequeños hacen que sea más difícil rechazar la hipótesis nula porque conducen a desviaciones estándar y errores estándar más grandes. También existe el riesgo de sesgo. Un ensayo pequeño que alcanzó significación (justificado o no) probablemente se publicará. [9].

La pregunta principal de la investigación que se aborda es: PI: “¿Para qué aplicamos teledetección en calidad de agua?”, y las subpreguntas de investigación:

SPI: “¿Qué nos indica la presencia de algas tóxicas con respecto a la calidad del agua?”

SP2: “¿Qué nos indica la presencia de microplásticos con respecto a la calidad del agua?”

SP3: “¿Qué índices se pueden generar a través de la teledetección de microplásticos o algas en los cursos de agua?”.

Definido como criterio de exclusión la antigüedad de publicación no mayor a tres años, los artículos anteriores a 2018 fueron excluidos y como criterio de selección de estudios se pondera la respuesta a 10(diez) preguntas definidas en una tabla con respuesta posible yes-no y parcial que mediante la siguiente operación matemática = $\text{CONTAR.SI}(\text{B12:K12};\text{"Yes"})+(\text{CONTAR.SI}(\text{B12:K12};\text{"Partial"}))/2$ otorga un puntaje y se eligen los estudios que obtuvieron valores mayores o iguales a 4(cuatro).

RESULTADOS OBTENIDOS

De 317.893 (trescientos diecisiete mil ochocientos noventa y tres) estudios encontrados. La RSL permitió diferenciar 33 (treinta y tres) documentos en inglés y español mediante la cadena de búsqueda, para luego de analizar si los estudios cumplían además de las preguntas y subpreguntas de investigación, pautas prefijadas como los criterios de inclusión y exclusión, y la valoración de calidad, se obtuvo un total de 7(siete) que son base para el desarrollo de éste artículo.

A partir de esos datos, se calculó el porcentaje de artículos totales publicados como resultado de una cadena de búsqueda. En el repositorio IEEE se encontraron 572 estudios que representan el 0,2% de artículos totales publicados, en ACM se encontraron 281.161 estudios que presentan el 88,4%, en Google Scholar se encontraron 5.770 estudios que representan el 1,8%, en UNLP se encontraron 14.785 estudios que representan el 4,7% y en UNSA se encontraron 11.028 estudios que representan el 3,5%.

Ponderando luego la información de los resúmenes con un puntaje, obtenido valorando las respuestas a las 10 preguntas. Se decide como elegibles los documentos que obtengan

4 (cuatro) o más puntos y esos fueron los finalmente citados para lectura y análisis.

	1. ¿Los estudios mencionan ODS?	2. ¿El estudio está diseñado para alcanzar dichos objetivos?	3. ¿Responde todas las preguntas de investigación adecuadamente?	4. ¿El estudio relaciona calidad de agua con microplásticos?	5. ¿El estudio relaciona calidad de agua con algas tóxicas?	6. ¿El estudio relaciona calidad de agua con sólidos en suspensión?	7. En el estudio se menciona extracción de datos de imágenes satelitales exportables?	8. ¿En el estudio se menciona extracción de datos de drones?	9. ¿En el estudio se presentan índices de calidad de agua?	10. ¿El estudio refleja análisis de imágenes satelitales procesadas con software libre?	Puntaje
Estudio múltiple	Yes	Partial	-	-	-	Yes	Yes	-	Yes	Yes	5,5
Estimación de la	Partial	Partial	-	-	-	Yes	Yes	-	-	Yes	4
Revolución de la	Yes	Partial	-	-	-	Partial	Yes	-	Yes	Yes	5
Análisis del esta-	Yes	Partial	-	-	-	-	Yes	-	Yes	-	3,5
USO DE IMAGE.	Yes	Partial	-	-	-	Yes	Yes	-	Yes	Yes	5,5
Optimización	-	-	Partial	-	-	-	Yes	-	Yes	Yes	3,5
Evaluación de	-	Partial	-	-	-	-	Yes	-	Yes	-	2,5
Investigación	-	-	Partial	Partial	Yes	Yes	Yes	-	Yes	Yes	6
Un algoritmo	-	-	Partial	Partial	Partial	Partial	Yes	-	Yes	Yes	5
Evaluación de	-	-	Partial	Partial	Partial	Partial	Yes	-	Yes	Yes	5
Mapa de la	-	Partial	Partial	Partial	Partial	Partial	Yes	-	Partial	Yes	4,5

Fig. 1. Ponderación y asignación de puntajes a los estudios.

En el análisis se obtuvieron datos acerca de las características de los estudios, Cada estudio fue evaluado respondiendo a las siguientes preguntas: 1-Los estudios mencionan ODS?, 2-El estudio está diseñado para alcanzar dichos objetivos?, 3-Responde todas las preguntas de investigación adecuadamente? 4-El estudio relaciona calidad de agua con microplásticos? 5-el estudio relaciona calidad de agua con algas tóxicas? 6-El estudio relaciona calidad de agua con sólidos en suspensión? 7-En el estudio se menciona extracción de datos de imágenes satelitales exportables? 8-En el estudio se menciona extracción de datos de drones? 9-En el estudio se presentan índices de calidad de agua? 10-El estudio refleja análisis de imágenes satelitales procesadas con software libre? Las respuestas contemplan tres posibilidades como mencionamos anteriormente, yes, no o parcial y los resultados hacen a un documento elegible cuando obtiene un puntaje superior o igual a 4 (cuatro). Como fin de este estudio e inicio de una nueva de etapa de investigación, se toman de los documentos relevados y ponderados con más puntajes las ideas que nutren las conclusiones, otorgando un enfoque de claridad sobre el tema de teledetección satelital y su relación y efecto en el análisis de datos obtenidos de las imágenes, terminología y aspectos básicos de temas relevantes.

Con respecto a la aplicación de la teledetección en calidad de agua: Los estudios analizados permiten comprender y dimensionar la aplicación de la teledetección en la calidad del agua. Con la información obtenida a partir de las imágenes satelitales, se pueden analizar grandes superficies de la tierra, disminuyendo tiempos y costos operacionales. Esto incluye la

posibilidad de investigar las cubiertas de agua, desde cuerpos pequeños hasta grandes masas oceánicas, considerándose como una alternativa eficaz para el estudio de dicho recurso natural. Este avance ha permitido un control y un conocimiento más ajustado de las condiciones atmosféricas, disminuyendo graves catástrofes naturales [10].

El análisis de los cursos de agua a través de teledetección se logra mediante “la interacción del flujo energético de los sensores con la superficie de la tierra” [7] hecho que recibe el nombre de radiación electromagnética. Esta interacción posibilita que “la adquisición de información por los sensores puede ser por reflexión, por emisión, y por emisión-reflexión” [11]. La adquisición de información por medio de los sensores, genera imágenes que pueden ser interpretadas mediante técnicas que permiten “analizar diversas variables biofísicas como la clorofila-a (Chl-a) y los sólidos totales en suspensión (SS), los cuales son de importancia para la calidad del agua” [12].

El avance en las técnicas de teledetección está facilitando este tipo de estudios debido a la mayor disponibilidad de imágenes y al gran desarrollo de nuevas tecnologías. Con estos métodos es posible obtener grandes cantidades de información con una resolución temporal, radiométrica y espacial elevadas, a un coste menor que con los métodos convencionales in situ [13]

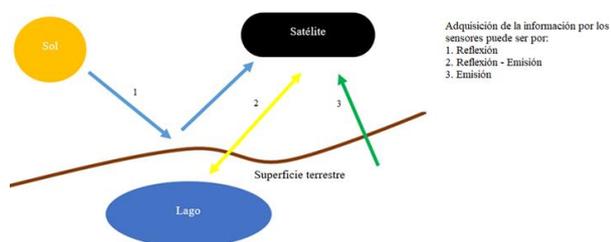


Fig. 2. Esquema de las formas de percepción remota. Adaptado de [8].

Se puede observar que las imágenes producto de la interacción del flujo energético de los sensores con la superficie de la tierra brindan información de grandes áreas a un menor coste, además, la posibilidad de monitorearlas

frecuentemente debido a que los satélites proporcionan datos que se encuentran en la web de forma gratuita. Calidad de agua: Calidad de agua se refiere al conjunto de parámetros indicadores del estado del agua para ser usada con ciertos propósitos. Es el grupo de concentraciones, especificaciones, sustancias orgánicas e inorgánicas y la composición de la biota encontrada en el cuerpo de agua analizado, cabe destacar que la calidad de agua se ve afectada cuando el agua sufre cambios que afectan su uso real o propósito [11].

Entre los grupos de clasificación de parámetros de calidad, los parámetros físicos responden a los sentidos del tacto, olor y sabor, los parámetros químicos están relacionados con la solvencia del agua y los parámetros biológicos están asociados a la calidad del medio acuático, y se basan en los organismos que lo habitan [11].

FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo se compone de un equipo interdisciplinario, 2 integrantes del equipo se recibieron de Magister en TI (UNaM-UNNE), 1 integrante está realizando su tesis de grado (UNaM), una integrante es doctorando del Doctorado en Informática (UNaM-UNNE-UTN (FRResistencia), desde donde sus integrantes contribuyen a la generación de conocimiento en áreas de ciencia de datos, gestión de calidad del agua, e imágenes satelitales.

CONCLUSIONES

Cuando los parámetros de calidad del agua no son adecuados para el uso real, podemos hablar de un deterioro. El deterioro de la calidad de las aguas superficiales se debe a la presencia de diversos tipos de contaminantes procedentes de actividades humanas como la agricultura, la industria, la construcción, la deforestación, etc. Así pues, la presencia de diversos contaminantes en las masas de agua puede conducir al deterioro tanto de la calidad

de las aguas superficiales como de la vida acuática [14].

Análisis in situ vs Análisis a través de teledetección: El avance de la tecnología permitió la existencia de otra forma para analizar la calidad del agua mediante técnicas de teledetección. A diferencia de los análisis in situ, “que sólo pueden representar estimaciones puntuales de la calidad del agua en un tiempo y espacio determinados. Algunas limitaciones de este método consisten en que el muestreo y las mediciones requieren mano de obra, tiempo y acarrear grandes gastos; el estudio de grandes áreas es casi imposible y a esto se suma variaciones espaciales y temporales y de las tendencias que son difíciles de seguir; también la exactitud y la precisión de los datos in situ recogidos pueden ser cuestionables [15].

BIBLIOGRAFÍA

- [1] Agua y saneamiento, <https://www.un.org/sustainabledevelopment/es/water-and-sanitation/>
- [2] Greenhalgh T. How to read a paper? Papers that summarize others papers (systematic reviews and metaanalyses). *BMJ* 1997; 315: 672-5.
- [3] J. Gopalakrishnan , P Ganeshkumar Systematic Reviews and Meta-analysis: Understanding the Best Evidence in Primary Healthcare, PMID: 24479036, PMCID: PMC3894019, DOI: 10.4103/2249-4863.109934
- [4] Glass, GV Primaria, secundaria y metaanálisis de la investigación. *Investigador educativo*, 5, 3-8. <https://doi.org/10.3102/0013189X005010033>, 1976
- [5] Xian Liu, Métodos y aplicaciones de análisis de datos longitudinales , 2016
- [6] Hunter, John E., & Schmidt, Frank L. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Newbury Park, CA: Sage.1990.
- [7] Jiyuan Liu a,* , Mingliang Liu a,b , Hanqin Tian a,b , Dafang Zhuang a , Zengxiang Zhang c , Wen Zhang d , Xianming Tang a , Xiangzheng Deng, Spatial and temporal patterns of China’s cropland during 1990–2000: An analysis based on Landsat TM data, 2000
- [8] Roderick P. McDonald and Moon-Ho Ringo Ho’Principles and Practice in Reporting StructuralEquation AnalysesUniversity of Illinois at Urbana–Champaign 2002
- [9] Stern, J.M. and Simes, R.J “Publication Bias: Evidence of Delayed Publication in a Cohort Study of Clinical Research Projects”. *BMJ*, 13, 640-645. 1997 <http://dx.doi.org/10.1136/bmj.315.7109.640>
- [10] D. A. Gutiérrez Núñez, "Evaluación del uso de la teledetección para determinar parámetros de la calidad del agua en el embalse de la Planta Hidroeléctrica de Tacaes", Trabajo de grado, Universidad de Costa Rica, 2019.
- [11] M. Escobar Valdivia, "Identificación de regiones contaminadas en la superficie del lago villarrica con base en imágenes sentinel en el periodo 2017-2018", Trabajo de grado, Universidad de Concepción, 2019.
- [12] "Estudio multitemporal de calidad del agua del embalse de Sitjar (Castelló, España) utilizando imágenes Sentinel-2", *Revista de teledetección*, vol. 56, n.º 117-130, octubre de 2020.
- [13] D. Uribe Ospina, "Estimación de la contaminación causada por la minería en cuerpos de agua del bajo cauca a través de imágenes satelitales", Trabajo de grado, UNIVERSIDAD EIA, 2019.
- [14] A. P. Cafa, "Monitorización de la calidad del agua de los lagos mediante técnicas de observación satelital", trabajo de especialización, ITBA, 2021.
- [15] D. C. Rivera Ruiz, "Estimación de parámetros de calidad de agua en la laguna santa elena usando imágenes satelitales", tesis maestría, Universidad de Concepción, 2020.

Logística inteligente para la recolección dinámica de residuos

D. Pandolfi, J. Rasjido, A. Villagra, S. Orozco, M. Lopez A. Cantella, A.M. Villagra, M. Bilbao, Pérez D.

[dpandolfi,jrasjido,avillagra,sorozco,mlopez,acantella,arivillagra,mbilbao,dperez}@uaco.unpa.edu.ar](mailto:{dpandolfi,jrasjido,avillagra,sorozco,mlopez,acantella,arivillagra,mbilbao,dperez}@uaco.unpa.edu.ar)

¹Laboratorio de Tecnologías Emergentes (LabTEM)
Instituto de Tecnología Aplicada (ITA) - Unidad Académica Caleta Olivia
Universidad Nacional de la Patagonia Austral

RESUMEN

Los procesos de recolección ineficientes y asignación no conveniente de vehículos conllevan a un aumento de las emisiones de gases contaminantes y de efecto invernadero, y mayores costos de proceso. Así entonces, la optimización de las rutas de la casa al vertedero para la recolección de residuos tecnológicos, producen beneficios tanto económicos como ambientales.

El problema de recolección de residuos tecnológicos es un problema derivado del Problema de Enrutamiento de Vehículos cuyo objetivo principal es la minimización de los kilómetros recorridos para los distintos escenarios dinámicos de recolección de residuos de manera resiliente, sustentable e inteligente. También, se analizará funciones derivadas para decidir el tipo de vehículo y cantidad de puntos de recolección. El coste vehicular, la emisión de gases efecto invernadero (CO₂), horas trabajadas, y consumo de combustible (diésel).

Adicionalmente, se aplicará al desarrollo de prototipos de aplicaciones móviles para una demanda dinámica de recolección de residuos y gestión de contenedores inteligentes.

Palabras clave: Ciudades Inteligentes, Residuos Tecnológicos, Problema de enrutamiento de vehículos, Inteligencia Computacional.

CONTEXTO

La línea de trabajo se lleva a cabo en el Laboratorio de Tecnologías Emergentes (LabTEM), Instituto de Tecnología Aplicada (ITA) de la Unidad Académica Caleta Olivia Universidad Nacional de la Patagonia Austral, en el marco del Proyecto de Investigación 29/B252 “Logística inteligente para la recolección dinámica de residuos”. Este proyecto se desarrolla en cooperación con el LIDIC de la UNSL, y el Grupo NEO de la UMA (España).

1. INTRODUCCIÓN

La logística verde (GL, *green logistics*) se ha convertido en una tendencia en la gestión de la distribución de bienes y la recolección de productos al final de su vida útil. Este enfoque tiene como objetivo maximizar el valor económico y ambiental mediante el reciclaje y el control de emisiones, GL contribuye al desarrollo sostenible y resiliente de la industria, pero también requiere un esquema de transporte más completo cuando se realizan servicios de logística (Lin et al. 2014 (a) (b)).

Los equipos eléctricos y electrónicos (EEE) son uno de los grupos más importantes de materiales que contienen residuos que son fáciles de reciclar. Sin embargo, contienen muchas sustancias que son tóxicas y potencialmente peligrosas para el medio ambiente y la salud humana (Oguchi et al., 2013). Los residuos electrónicos se conocen como WEEE (Residuos de

equipos eléctricos y electrónicos) o residuos electrónicos (e-Waste). La tasa de desechos electrónicos está creciendo en la actualidad, especialmente en los países desarrollados, donde los mercados están saturados con grandes cantidades de nuevos productos electrónicos (Oguchi et al., 2012). En 2013, el StEPI (Iniciativa para resolver el problema de los residuos electrónicos) informa que la generación global de residuos electrónicos se estimó en alrededor de 44,7 millones de toneladas en 2016, una cifra que se prevé que aumente alrededor de 50,7 millones de toneladas en 2020. Europa (incluida la Federación de Rusia) ha generado la segunda mayor cantidad de residuos electrónicos per cápita (16,6 kg por habitante) después de Oceanía (17 kg por habitante) en 2016 (Baldé et al., 2017).

Un método de recolección dinámica de e-Waste, propone un sistema de recolección a pedido o demanda (Nowakowski et al., 2017). Este método de recolección consiste en una notificación previa por parte de un residente o una compañía para recogida de equipos de desecho. Un sistema de soporte de recolección móvil puede ofrecer grandes posibilidades para la reducción de costos de la recolección con una secuencia de recogida de equipos adecuadamente preparada. Para respaldar tal método de recolección, se debe aplicar un algoritmo con capacidades de planificación inteligente y un sistema de información eficiente para la eliminación de desechos de los residentes urbanos. (Nowakowski et al., 2018, 2020).

El Problema de enrutamiento de vehículos (VRP) Dantzig y Ramsar (1959), es uno de los problemas más conocidos y desafiantes en la programación lineal entera y es parte de los problemas llamados NP-Hard (Lenstra y Kan, 1981), que son los problemas que no se pueden resolver en un tiempo polinomial. El tiempo y esfuerzo computacional requerido para resolver este problema aumenta exponencialmente respecto al tamaño del problema. Por esta razón se han ideado algoritmos que no garantizan optima-

lidad, pero que logran entregar buenas soluciones a estos problemas difíciles de resolver como lo son las heurísticas y metaheurísticas.

Con el objetivo de establecer formalmente el proceso de recolección seleccionado y el método de enrutamiento, puede considerarse lo propuesto por Dorigo et al. 1996 para el problema de ruteo de vehículos con capacidad limitada y el algoritmo de Optimización con colonias de hormigas (en inglés *Ant Colony Optimization*, ACO), tomando como base este estudio se propone un método para optimizar las rutas de recolección y operarlas dinámicamente. Las distintas estrategias de recolección de e-waste pueden asociarse al conocido problema de enrutamiento de vehículos. En general, VRP consiste en asignar una cantidad de vehículos homogéneos a varios clientes, donde cada cliente tiene una cierta ubicación y demanda de productos (homogéneos). El objetivo de la optimización es entregar los bienes demandados a todos los clientes al tiempo que minimiza la suma de los costos (longitudes) de las rutas de los vehículos.

La complejidad de muchos problemas optimización discreta del mundo real está asociada con grandes espacios de búsqueda, demandas de rendimiento de tiempo real y ambientes dinámicos que no pueden ser resueltos por métodos exactos en tiempo razonable. Esto ha promovido al desarrollo de nuevos enfoques conocidos como metaheurísticas. Estas, son métodos que integran de diversas maneras, procedimientos de mejora local y estrategias de alto nivel para crear un proceso capaz de escapar de óptimos locales y realizar una búsqueda robusta en el espacio del problema. En particular, las soluciones a problemas de planificación de rutas reales requieren enfoques específicos que puedan manejar tanto la complejidad intrínseca como la incertidumbre de configuración. Las metaheurísticas son técnicas algorítmicas cuyo propósito general es ofrecer una estrategia de búsqueda global para explorar un espacio de soluciones. Las principales ventajas de las metaheurísticas son su adaptabilidad a las características del problema y su conocimiento del

mismo problema para encontrar una solución. Los procedimientos de resolución de problemas son capaces de abordar problemas complejos de optimización, incluidos aquellos con información imprecisa, evaluar posibles alternativas y determinar de manera eficiente la solución preferida.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Esta línea de trabajo se enfoca en la generación de rutas recolección que permitan minimizar ciertos factores económicos y ambientales y maximizar los beneficios referidos a al reciclaje y tratamiento de WEEE. De los objetivos económicos podemos mencionar: minimizar el tiempo de recolección, kilómetros recorridos, maximizar el ahorro de combustible en los vehículos, minimizar la cantidad de vehículos, todo lo cual llevaría a obtener menores costos. Un objetivo económico adicional está dado por la valoración de la recuperación las materias primas tales como material ferroso, cobre, y metales preciosos, etc. De los beneficios ambientales podemos mencionar: la disminución en la emisión de gases efecto invernadero (CO_2 , CH_4 y N_2O) y los gases de contaminación ambiental (CO , SO_2 , PM , NO_x). Tanto los objetivos económicos como ambientales promueven una mejor calidad de servicio e imagen, para las empresas y los gobiernos locales.

3. RESULTADOS OBTENIDOS/ESPERADOS

El problema de recolección de e-waste estacionaria se lleva a cabo en puntos de recolección específicos. Dichos puntos, están ubicados en lugares populares o en puntos centroides donde se hace mínimo la distancia entre cada residencia y el punto limpio de recolección. La elección de la ubicación de dichos centros y la elección de un recorrido óptimo representa distintos problemas de decisión.

Para la recolección dinámica por demanda o en puntos móviles de recolección debe incluirse dos aspectos a desarrollar, aplicaciones móviles para transaccionar la demanda y contenedores inteligentes que indiquen ubicaciones y estado de llenado de estos.

El estudio desarrollado corresponde a la de recolección en puntos fijos de pequeños e-waste. Para la recolección estacionaria se establecieron cuatro escenarios distintos de 75, 105, 150 y 200 puntos de recolección de e-waste, ubicados en lugares con disponibilidad para su gestión. La elección de la ubicación de dichos centros fue establecida de forma aleatoria, en distintos centros educativos o gubernamentales de las localidades de Caleta Olivia y Comodoro Rivadavia.

Los contenedores son todos de igual tamaño con un volumen de 0,225 m³. Los vehículos de recolección poseen distintas capacidades y características. El primero de los vehículos soporta una carga de hasta 1,5 toneladas y cubicaje de 16 m³, el coste promedio por km recorrido es de 1,1526 euros, tiene un consumo de 0,30 litros de gasoil por km recorrido, tiene una capacidad de carga máxima de 71 contenedores y su carga y descarga es de tipo manual requiriendo 6 y 2 minutos respectivamente

Como objetivo general se espera modelar, resolver e implementar un amplio y variado conjunto de servicios inteligentes de la ciudad y producir un mayor impacto en Ciencia e Industria.

Como objetivos específicos se pretende:

- Contribuir a la disminución de la contaminación ambiental y el efecto invernadero a través de la aplicación de técnicas de inteligencia artificial para problemas de logística verde y minería inversa.
- Optimizar el proceso de recolección de residuos para escenarios dinámicos y estimar diferentes variables de decisión (costos, ambientales y recuperación de materiales).
- Construir prototipos de contenedores inteligentes para de determinación temprana de la oportunidad de recolección.
- Promover y contribuir el desarrollo de la economía circular a través de reparación, reciclaje y recuperación de

minerales. e) Promover la concientización sobre el cuidado del medioambiente y la deposición final responsable de la tecnología. f) En lo académico e institucional se espera profundizar la formación del equipo de investigadores participantes y becarios y alumnos participantes

Actualmente se han generado cuatro instancias de pruebas para para 75, 105, 150 y 210 locaciones de recogida de WEEE de cuatro localidades (Caleta Olivia, Comodoro Rivadavia, Trelew y Río Gallegos). Las instancias se han probado con dos algoritmos de inteligencia computacional (ACO y cGA) los resultados preliminares fueron publicados en Pandolfi et al. 2019 y Perez et al 2020.

Para elegir los vehículos más adecuados para la colección, analizamos las siguientes variables: total de km recorridos, Costo total de uso vehículo, Total de Combustible consumido, Total de CO2 emitido y Tiempo Total de horas de Trabajo.

En trabajos futuros analizaremos en distintas ciudades y se propondrán nuevas funciones de optimización teniendo en cuenta las variables consideradas en este estudio.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo se encuentra formado por cuatro investigadores con distintos niveles de posgrado, dos Doctores en Ciencias de la Computación, un Magister en Ciencias de la Computación, tres Ingenieros en Sistemas, tres estudiantes de la Carrera Ingeniería en Sistemas.

Esta línea de investigación proporcionará un marco propicio para la iniciación y/o finalización de estudios de posgrado de los integrantes docentes. De igual forma, será un ámbito adecuado para la realización de tesis de grado. En ese sentido, dos integrantes de este proyecto de investigación está desarrollando su Tesis de Maestría en temáticas afines. Además, se cuenta con dos becarios alumno de la carrera de Ingeniería en Sistemas.

BALDÉ C.P., FORTI V., GRAY V., KUEHR R., STEGMANN, P. (2017). The global e-waste monitor – 2017. Bonn, Geneva, Vienna: United Nations University, International Telecommunication Union, International Solid Waste Association.

CARBALLO L., VILLAGRA A., ERRECALDE M., Movilidad inteligente: reducción de emisión de gases. Revista de Informes Científicos y Técnicos de la Universidad Nacional de la Patagonia Austral. vol.11 n°2. Pág.53-69. ISSN 1852-4516 (2019).

DANTZIG G.B., RAMSER J., (1959). The truck dispatching problem, Manage. Sci. 6 (1) (1959) 80–91.

DORIGO M., MANIEZZO, COLORNI A., (1996). The ant system: optimization by a colony of cooperating agents, IEEE Transactions on Systems, Man, and Cybernetics Part B26 29–41.

LENSTRA J.K., KAN A.R., (1981). The complexity of vehicle routing and scheduling problems, Networks 11 (1981) 221–227.

LIN C., CHOY K.L., HO G.T.S., Ng T.W. (a), A Genetic Algorithm-based optimization model for supporting greentransportation operations, Expert Systems with Applications, Volume 41, Issue 7, 2014, Pages 3284-3296, ISSN 0957-4174.

LIN C., CHOY K.L., HO G.T.S., CHUNG S.H., LAM H.Y. (b), Survey of Green Vehicle Routing Problem: Past and future trends, Expert Systems with Applications, Vo-lume 41, Issue 4, Part 1, 2014, Pages 1118-1138, ISSN 0957-4174.

LOPEZ M., PANDOLFI D., VILLAGRA A.; Heurística y metaheurísticas en la recolección de residuos tecnológicos; Revista de Informes Científicos y Técnicos de la Universidad Nacional de la Patagonia Australn 2021.

BIBLIOGRAFÍA

NOWAKOWSKI P., KRÓL A., MRÓWCZNSKA B. (2017), Supporting mobile WEEE collection on demand: A method for multi-criteria vehicle routing, loading and cost optimisation, *Waste Management*, Volume 69, 2017, Pages 377-392, ISSN 0956-053X.

NOWAKOWSKI P., SZWARC K., BORYCZKA U., (2018) Vehicle route planning in e-waste mobile collection on demand supported by artificial intelligence algorithms, *Transportation Research Part D: Transport and Environment*, Volume 63, 2018, Pages 1-22, ISSN 1361-9209.

NOWAKOWSKI, P., SZWARC, K., & BORYCZKA, U. (2020). Combining an artificial intelligence algorithm and a novel vehicle for sustainable e-waste collection. *Science of The Total Environment*, 730, 138726.

OGUCHI M., SAKANAKURA H., TERAZONO A. (2012). Toxic metals in WEEE: characterization and substance flow análisis in waste treatment processes. *Science of The Total Environment*, Volumes 463–464, 2013, Pages 1124-1132, ISSN 0048-9697.

OGUCHI M., SAKANAKURA H., TERAZONO A. (2012). Toxic metals in WEEE: characterization and substance flow análisis in waste treatment processes. *Science of The Total Environment*, Volumes 463–464, 2013, Pages 1124-1132, ISSN 0048-9697.

PEREZ D., PANDOLFI D., VILLAGRA A.; Recolección de Residuos Tecnológicos aplicando Metaheurísticas. *Revista de Informes Científicos y Técnicos de la Universidad Nacional de la Patagonia Austral*. Vol. 13 Núm. 1 (2021). Pag. 54-76.

PANDOLFI D., VILLAGRA A., RASJIDO J., LEGUIZAMON G. Following the green footprint of technological waste: A smart and sustainable collection. XVIII RPIC - Reunión de Trabajo en Procesamiento de la Información y

Control. Universidad Nacional del Sur. Septiembre 2019. Pág. 287-292. ISBN 978-987-1648-44°-3

VALDEZ J., PANDOLFI D., VILLAGRA A. Red de sensores inteligentes para adquisición de datos de una planta de desalinización de agua. *Revista de Informes Científicos y Técnicos de la Universidad Nacional de la Patagonia Austral*. vol. 10 n°3. Pág. 83-95. ISSN 1852-4516 (2018).

VILLAGRA A., PANDOLFI D., RASJIDO J., MERCADO V. Hibridación de Metaheurísticas aplicadas al problema de ruteo de vehículos con capacidad uniforme. IX Seminario Euro Latinoamericano de Sistemas de Ingeniería. Venezuela. Noviembre 2013. ISBN: 978-980-7630-00-9.

VILLAGRA A., ALBA E., LUQUE G. A Better Understanding on Traffic Light Scheduling: New Cellular GAs and New In-depth Analysis of Solutions. *Journal of Computer Science*. 41 (2020) 101085. ISSN: 1877-7503 <https://doi.org/10.1016/j.jocs.2020.101085>

Procesamiento Automático de Imágenes Médicas

García Mario Alejandro ✉, Gramática Martín Nicolás

Universidad Tecnológica Nacional Facultad Regional Córdoba (UTN FRC)

mgarcia@frc.utn.edu.ar

RESUMEN

La Inteligencia Artificial ha logrado grandes avances en la medicina, pero la mayoría de estos avances no llegan a la práctica clínica. En esta línea de investigación, a través de un proyecto que le da inicio, se enfoca en los desafíos para lograr la aplicación de la Inteligencia Artificial en la práctica clínica. En primer lugar se analizan los obstáculos conocidos y soluciones propuestas, para trabajar después sobre un caso concreto de Patología Digital. Se planea entrenar modelos de aprendizaje profundo con una imagen WSI completa combinando técnicas de compresión de imágenes y *gradient checkpointing*.

1. CONTEXTO

El proyecto UTN8436, “Procesamiento Automático de Imágenes Médicas”, perteneciente al Grupo de Inteligencia Artificial (GIA) de la Universidad Tecnológica Nacional Facultad Regional Córdoba, pretende ser el inicio una línea de investigación donde se concentren trabajos de Inteligencia Artificial aplicada al reconocimiento de patrones en imágenes médicas que tengan como foco la implementación real en la práctica clínica.

2. INTRODUCCIÓN

La inteligencia artificial (IA) y el aprendizaje profundo (AP) han logrado en los últimos años grandes avances para su aplicación en medicina, y más específicamente, en las tareas

relacionadas con el diagnóstico por imágenes. Por ejemplo, Kather *et al.* en (Kather, 2019) predicen la inestabilidad de microsatélites (MSI por su sigla en inglés) en cáncer gastrointestinal directamente sobre imágenes de microscopio y Sarker *et al.* en (Sarker, 2020) detectan infecciones por COVID-19 en radiografías de tórax.

1.1. Diagnóstico automático sobre imágenes

En el diagnóstico automático sobre imágenes, las técnicas de IA se utilizan principalmente en las tareas de clasificación de imágenes/exámenes, clasificación de objetos o lesiones, localización de órganos, regiones y puntos de referencia, detección de objetos o lesiones, segmentación de órganos o subestructuras, segmentación de lesiones, registración, recuperación de imágenes basada en contenido, generación y mejora de imágenes y combinación de reportes con datos de imágenes. Más detalles sobre aplicaciones se pueden ver en (Litjens, 2017).

La mayoría de las tareas mencionadas se pueden llevar a cabo sobre imágenes de áreas grandes o imágenes de microscopía. Las imágenes de microscopía pertenecen al área de la patología y microscopía digital, muchas veces llamada simplemente Patología Digital (PD). En PD, la imagen se administra con una tecnología especial llamada microscopía virtual o, más comúnmente, *whole-slide images* (WSI). Una imagen WSI es una foto de tejidos en la escala de los gigapíxeles que

se almacena mediante una estructura jerárquica. Un panorama más amplio sobre el tema se puede ver en (Pantanowitz, 2018) y (Kobayashi, 2021).

1.2. Dificultades en la implementación clínica

Si bien las técnicas de IA aplicadas al diagnóstico por imágenes han demostrado un gran potencial para mejorar los diagnósticos médicos, muy pocos algoritmos han alcanzado la implementación clínica (van der Laak, 2021). En (Kather, 2019) se puede ver un ejemplo del problema. Kather *et al.* reconocen MSI sobre imágenes WSI teñidas con hematoxilina-eosina usando dos redes neuronales ResNet-18. El rendimiento se evaluó con el área bajo la curva (AUC) ROC y el área obtenida fue 0.84 sobre la base de datos de validación, pero el mismo modelo logró $AUC < 0.69$ cuando se evaluó con una base de datos japonesa, probablemente porque la base de datos de entrenamiento estaba compuesta por un 80% de no asiáticos. Esto indica la necesidad de datos de entrenamiento multicéntricos para obtener modelos de clasificación más generales. Además de las diferencias entre pacientes de distintas regiones, las muestras de distintos laboratorios tienen gran variabilidad en la tinción, la calidad de la imagen, las características de escaneo y la preparación de los tejidos.

3. DESAFÍOS

Los desafíos que aún deben abordarse para que la IA pueda alcanzar un valor clínico han sido tratados en numerosos trabajos (Colling, 2019), (Acs, 2020), (van der Laak, 2021), (Yoshida, 2021), (Kobayashi, 2021), (Reinke, 2021). A continuación se hace un breve resumen.

Conjuntos de datos a gran escala. Gracias a los prometedores resultados de las primeras aplicaciones de IA, el tamaño de los conjuntos de datos ha aumentado, lo que ha llevado a un número cada vez mayor de esfuerzos multicéntricos entre distintos laboratorios. Sin embargo, aunque recopilar una gran cantidad de WSI es una tarea manejable, el etiquetado sigue siendo un obstáculo para la escala de los algoritmos de AP. El etiquetado puede significar tanto la anotación manual de regiones de la imagen como las anotaciones de historia clínica. La adquisición de anotaciones manuales de imágenes es una tarea tediosa que requiere experiencia en el dominio y generalmente la realizan patólogos. Por el contrario, las anotaciones clínicas requieren acceso a informes de patología y registros electrónicos de pacientes, ya sea de un hospital o de un registro regional o nacional. El etiquetado clínico de WSI tiende a ser más fácil de lograr que la anotación manual y ha dado lugar a grandes conjuntos de datos en varios estudios. Aun así, se piensa que la construcción de modelos de IA utilizando sólo anotaciones clínicas no será posible ni eficiente para todas las aplicaciones en diagnóstico por imágenes. Por lo tanto, el etiquetado manual seguiría siendo necesario, lo que requerirá el desarrollo de técnicas para facilitar la producción eficiente de estas anotaciones. Se han propuesto varios enfoques:

- El enfoque más simple (trivial) es aumentar el número de etiquetadores. Este enfoque tiene la ventaja de garantizar anotaciones de alta calidad, pero es muy costoso.
- Técnicas de tinción como la inmunohistoquímica (IHC), en las que se pueden usar anticuerpos para atacar tipos específicos de tejido o células.

- Otra técnica útil es la re-tinción. Esta proporciona una alternativa a los cortes seriados, en la que se puede teñir la misma muestra con H&E e IHC y alinear las muestras digitalizadas mediante algoritmos especiales..

Aprendizaje débilmente supervisado. Es otro enfoque para reducir la carga del etiquetado. En el contexto de la segmentación de imágenes, la supervisión débil puede presentarse en forma de anotaciones manuales escasas (por ejemplo, el etiquetado de pequeñas regiones utilizando puntos o garabatos, en contraposición a la supervisión completa en la que todos los píxeles son etiquetado). La aplicación de este enfoque sobre imágenes WSI es un gran desafío. Por un lado, se podría utilizar solo una etiqueta para la imagen completa, como por ejemplo, la presencia de cáncer. Por otro lado, como el tumor no está presente en todos los *slides*, no hay una referencia para el entrenamiento del modelo con cada imagen de bajo nivel. El entrenamiento extremo a extremo (WSI completo) en este caso es imposible porque una imagen WSI puede fácilmente requerir decenas de gigabytes de memoria de GPU. Algunos enfoques de solución son:

- Asumir que todas las partes de la imagen WSI contienen información correlacionada con la etiqueta. Este enfoque no funciona para objetos muy pequeños o poco frecuentes.
- Reducir el tamaño de las imágenes WSI hasta que puedan ser procesadas. Estos enfoques se basan en la compresión de imágenes WSI utilizando redes neuronales.
- Otros enfoques utilizan *gradient checkpointing* para reducir la cantidad de memoria utilizada por el método

backpropagation durante el entrenamiento de las redes neuronales.

Generalizabilidad de los algoritmos de IA a la práctica clínica. Si bien el tamaño de las bases de datos crecido sustancialmente, muchas todavía carecen de una variabilidad cercana a la real. Estas variaciones pueden generar sesgos en los diagnósticos.

- Una idea es introducir variabilidad artificial con aumentación de datos (*data augmentation*).
- Un enfoque alternativo (o complementario) es la normalización de las imágenes a un estándar común bajo la hipótesis de que la variabilidad se puede eliminar.

Un problema adicional es que los modelos de IA solo reconocen los patrones para los que fueron entrenados. Un sistema que reemplace a un médico en el análisis de una imagen donde se busca indicios de una enfermedad particular, pasaría por alto pistas sobre otro problema. Una alternativa es desarrollar modelos con una salida específica adicional para indicar un nivel de duda.

Validación. Los resultados de los sistemas de IA son validados habitualmente con dos juegos de datos independientes del entrenamiento (validación y test), aunque frecuentemente pertenecen a la misma base de datos. En medicina esto podría ser insuficiente, incluso para los casos donde se valida con otra base de datos (validación externa). Algunos autores sugieren que los modelos deberían ser sometidos a procesos de validación similares a los que se utilizan con drogas u otros procedimientos, donde se analizan los efectos a largo plazo.

Un tema central en la validación es la elección de las métricas de desempeño de los modelos.

(Reinke, 2021) es un documento actualizado dinámicamente por investigadores que ilustra las limitaciones de las métricas que se aplican frecuentemente en el campo del análisis de imágenes.

IA explicable. En medicina es importante conocer los fundamentos de cada decisión, pero las soluciones basadas en redes neuronales (incluyendo AP) pueden verse como una caja negra. La explicabilidad de la IA es un tema de investigación activo que excede a las aplicaciones en medicina. Una revisión de los enfoques para lograr explicaciones se puede ver en (Guidotti, 2018). Existen numerosos artículos que tratan el tema. En (Angelov, 2020) por ejemplo, se muestra un enfoque para explicar las salidas de modelos de AP y en (Lee, 2019) se puede ver una red neuronal “explicable” para la detección de hemorragias intracraneales agudas.

4. OBJETIVOS

El objetivo principal del proyecto UTN8436 es contribuir a las transformaciones necesarias para que los sistemas de reconocimiento de patrones sobre imágenes médicas, que han logrado resultados satisfactorios en experimentos de investigación, se implementen en la práctica clínica.

Objetivos específicos:

- Detectar y validar los principales factores que dificultan la transición de sistemas entre los ambientes de investigación y de medicina aplicada.
- Estudiar en detalle las soluciones propuestas por la comunidad científica para cada una de las causas del punto anterior.
- Reproducir resultados de experimentos de diagnóstico

publicados, principalmente sobre WSI, y aplicar técnicas propuestas, modificaciones de las mismas o nuevos enfoques para la solución de alguno de los problemas particulares detectados.

- Realizar un seguimiento de los estándares de evaluación para técnicas de IA en medicina (tal como se explicó en la sección anterior, se encuentran en evolución) y evaluar tanto los resultados de los experimentos como los métodos propuestos según estos estándares.

5. DESARROLLO

Después de un análisis profundo de los desafíos para la aplicación clínica de la IA, el proyecto UTN8436 se enfocará principalmente en un caso concreto del área de PD con aprendizaje débilmente supervisado. Se planea entrenar modelos de aprendizaje profundo con una imagen WSI completa combinando técnicas de compresión de imágenes y *gradient checkpointing*.

6. FORMACIÓN DE RECURSOS HUMANOS

Además de la formación de los recursos humanos (docentes/investigadores y becarios) que integran el proyecto vigente, se busca albergar nuevos proyectos relacionados y promover el desarrollo tesis de todos los niveles en el tema.

Por otro lado, experiencia de esta línea de investigación enriquece la cátedra de Inteligencia Artificial.

REFERENCIAS

Acs, Balázs, Mattias Rantalainen, and Johan Hartman. "Artificial intelligence as the next

step towards precision pathology." *Journal of internal medicine* 288.1 (2020): 62-81.

Angelov, Plamen, and Eduardo Soares. "Towards explainable deep neural networks (xDNN)." *Neural Networks* 130 (2020): 185-194.

Colling, Richard, et al. "Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice." *The Journal of pathology* 249.2 (2019): 143-150.

Kather, Jakob Nikolas, et al. "Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer." *Nature medicine* 25.7 (2019): 1054-1056.

Kobayashi, Soma, Joel H. Saltz, and Vincent W. Yang. "State of machine and deep learning in histopathological applications in digestive diseases." *World Journal of Gastroenterology* 27.20 (2021): 2545.

Litjens, Geert, et al. "A survey on deep learning in medical image analysis." *Medical image analysis* 42 (2017): 60-88.

Guidotti, Riccardo, et al. "A survey of methods for explaining black box models." *ACM computing surveys (CSUR)* 51.5 (2018): 1-42.

Lee, Hyunkwang, et al. "An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets." *Nature Biomedical Engineering* 3.3 (2019): 173-182.

Pantanowitz, Liron, et al. "Twenty years of digital pathology: an overview of the road travelled, what is on the horizon, and the emergence of vendor-neutral archives." *Journal of pathology informatics* 9 (2018).

Reinke, Annika, et al. "Common limitations of image processing metrics: A picture story." *arXiv preprint arXiv:2104.05642* (2021).

Sarker, Md Mostafa Kamal, et al. "Web-based efficient dual attention networks to detect COVID-19 from X ray images." *Electronics Letters* 56.24 (2020): 1298-1301.

Turner, Oliver C., et al. "Mini Review: The Last Mile—Opportunities and Challenges for Machine Learning in Digital Toxicologic Pathology." *Toxicologic pathology* 49.4 (2021): 714-719.

van der Laak, Jeroen, Geert Litjens, and Francesco Ciampi. "Deep learning in histopathology: the path to the clinic." *Nature medicine* 27.5 (2021): 775-784.

Yoshida, Hiroshi, and Tomoharu Kiyuna. "Requirements for implementation of artificial intelligence in the practice of gastrointestinal pathology." *World Journal of Gastroenterology* 27.21 (2021): 2818.

Zhou, S. Kevin, et al. "A review of deep learning in medical imaging: Image traits, technology trends, case studies with progress highlights, and future promises." *Proceedings of the IEEE* vol 109, no. 5 (2021): 820-838.

Sistemas conversacionales aplicados a la gobernanza

F. Brandan¹, D. Pandolfi¹, A. Villagra¹, M. Errecalde^{1,2}
[fbrandan, dpandolfi, avillagra }@uaco.unpa.edu.ar](mailto:{fbrandan, dpandolfi, avillagra}@uaco.unpa.edu.ar), merreca@unsl.edu.ar

¹Laboratorio de Tecnologías Emergentes (LabTEm)
Instituto de Tecnología Aplicada (ITA) - Unidad Académica Caleta Olivia
Universidad Nacional de la Patagonia Austral

²Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)
Departamento de Informática - Universidad Nacional de San Luis

RESUMEN

Un chatbot es un sistema inteligente conversacional capaz de simular y procesar una conversación humana de forma escrita o hablada, permitiendo a sus usuarios interactuar con servicios digitales como si estuvieran tratando con otra persona real. Estos pueden ser tan sencillos como un pequeño programa que responde preguntas con una línea de texto, como un gran y sofisticado asistente digital que aprende y relaciona conceptos a partir de la información que recibe y procesa logrando altos niveles de personalización e “inteligencia”.

El chatbot propuesto será capaz de responder consultas y entregar formularios a los interesados en los cursos y trabajos disponibles de la oficina de empleo local en Caleta Olivia utilizando lenguaje natural ameno y sencillo de seguir, también tomando en cuenta posibles reconocimientos erróneos de intenciones y pudiendo aprender de ellos.

Palabras clave: Inteligencia; Comunicación; Redes Neuronales; NPL, Chatbot.

CONTEXTO

La línea de trabajo se lleva a cabo en el Laboratorio de Tecnologías Emergentes (LabTEm), Instituto de Tecnología Aplicada (ITA) de la Unidad Académica Caleta Olivia Universidad Nacional de la Patagonia Austral, en el marco del Proyecto de Investigación 29/B273 “Ciudades inteligentes y sostenibles: iniciativas

y desafíos”. Este proyecto se desarrolla en cooperación con el LIDIC de la UNSL, y el Grupo NEO de la UMA (España).

1. INTRODUCCIÓN

El objetivo principal de toda herramienta es facilitar el desempeño de una tarea; y podemos ver a la tecnología como un proceso o capacidad de transformar o combinar algo existente para construir algo nuevo. Por lo tanto, podemos decir que, la tecnología nos permite crear nuevas herramientas para resolver cuestiones del día a día.

El avance incesante de las nuevas tecnologías llevó al mundo a una nueva revolución del modo de vivir, amplificado por el contexto de pandemia durante los años 2020 y 2021. Este proceso de reestructuración ocurre en todos los campos, pero resulta en especial aparente en el entorno laboral y social.

Según estudios del INDEC (2020) en Argentina, 86 de cada 100 personas utilizan Internet y 88 de cada 100 emplean teléfono celular. En términos comparativos, respecto del mismo periodo del año anterior, se registra un incremento en el uso del Internet (5,6% más) y del teléfono celular (3,6% más). Estamos hablando de un país que cada vez utiliza más las nuevas TIC (Tecnologías de la Información y Comunicación), un entorno donde la eficiencia de comunicación debe

ser la mayor posible, la relación entre los recursos empleados y los resultados debe ser la mejor. La disponibilidad y calidad de las comunicaciones se transforma en un punto central a futuro. Con esto empezamos a ver una imagen clara sobre a dónde se dirige el foco de atención del desarrollo del país, se intenta llegar al concepto de ciudad inteligente.

Es importante definir concretamente qué es un chatbot o bot conversacional, porque en la actualidad la mayoría de las personas han interactuado con un bot de estas características sabiéndolo o no, y está tan integrado en el día a día que vale la pena ver a detalle a qué nos referimos ORACLE (2022).

Las tres funciones principales de las que está compuesto un chatbot son las siguientes:

- Agente de diálogo: Capaz de recibir los mensajes del usuario. El bot recibe una entrada textual u oral, que es analizada para generar una respuesta adecuada.
- Agente racional: El bot debe tener acceso a una base de datos con las respuestas adecuadas a las preguntas del usuario. Además, debe ser capaz de almacenar información propia del usuario y del contexto del mensaje, como nombre del usuario, fechas, género, y otros datos similares para dar una respuesta adecuada.
- Agente personificado: El lenguaje del bot debe ser lo más natural y real posible, por lo que se le puede darle una personalidad con el objetivo de crear una confianza con el usuario y dar una sensación de charla con otra persona. Comúnmente se les asigna nombres a estos asistentes, por ejemplo, Eliza, Alice, Padi, entre otros.

Las posibilidades de un chatbot residen en las decisiones del desarrollador y sus necesidades. Podrían resolver múltiples tipos de preguntas o problemas de un cliente, automatizar procesos de forma ilimitada, y al mismo tiempo reduciendo la necesidad de interacción humana

Courtaudon (2012). Se puede obtener una mejora de la satisfacción de los usuarios cuando se les ofrece la posibilidad de utilizar un chat donde pueden ser fácilmente guiados a donde deben dirigirse y resolver sus dudas sin la necesidad de ponerse en contacto con un empleado y espera la respuesta. Para poder realizar esto se debe profundizar en el concepto de “Procesamiento del Lenguaje Natural” (NLP de *Natural Language Processing*). Sus comienzos se dan en 1950 con el artículo publicado por Alan Turing *Computing Machinery and Intelligence* (Maquinaria de computación e inteligencia) donde discute la temática de la inteligencia artificial y se introduce el concepto de la Prueba de Turing.

A partir del desarrollo de la Web 2.0 o Web social (donde el usuario empieza a tomar un rol más activo en la comunicación y es capaz de entregar datos de la misma forma que los recibía en la Web estática) en los años 2000s se habilita la recolección de datos en grandes cantidades, lo que impulsa a que el aprendizaje automático sea mucho más viable. El enfoque se mueve a algoritmos con baja o sin supervisión que toman datos en bruto y saben elegir la información relevante para sus tareas.

Actualmente el enfoque principal de este tipo de proyectos se encuentra en el *Deep learning* (aprendizaje profundo), refiriéndose al uso de múltiples capas en una red. Se crean redes neuronales artificiales (ANN) inspiradas en el procesamiento de información y nodos de comunicación distribuidos de forma similar a los sistemas biológicos. Aunque a diferencia de un cerebro, las redes neuronales tienden a permanecer estáticas y simbólicas, mientras que el cerebro biológico es dinámico y análogo. El uso de esta técnica ayudó a lograr grandes logros en el reconocimiento y modelaje de lenguaje.

Los principales desafíos del NLP, YAN (2020) que se enfrentarán en este proyecto son la Comprensión del Lenguaje Natural (*Natural Language Understanding*) donde se aplica un preprocesamiento al texto escrito o hablado y

pueda ser interpretado por la máquina fácilmente. Y por otro lado la Generación del Lenguaje Natural (*Natural Language Generation*) donde la máquina transforma los datos estructurados respuesta y los transforma en lenguaje natural para comprensión del usuario.

Los Chatbots inteligentes se han convertido en la tecnología más importante para impulsar la automatización de las aplicaciones móviles y ofrecer resultados nunca vistos para empresas de todos los segmentos de actividad. La clave en la estrategia de marketing residirá en el nivel de innovación de los bots que cada empresa pueda incorporar en sus aplicaciones móviles Adiwardana et al. (2020).

Recientemente, los Chatbots se han vuelto más flexibles y útiles gracias la ubicuidad de los smartphones y los avances en *Machine Learning* (ML) y *Deep Learning* (DL). Una manera de clasificar los chatbots es ver como ellos interactúan con los usuarios, a) chatbots que tienen respuestas, un conjunto de respuestas fijas o limitadas, b) Bot que se basan en flujos conversacionales más complejos en términos de variabilidad de respuestas, y c) y Bots Open-ended utilizados principalmente en entretenimiento donde pueden conversar de varios tópicos con los usuarios. Para construir un sistema de diálogo se debe construir distintas componentes en el flujo de diálogo 1) Reconocimiento de Mensaje, 2) Comprensión del Lenguaje Natural, 3) Gestión del diálogo, gestión de tareas 4) generación de lenguaje natural, 5) generación de respuestas en textos Bunk et al. (2020), Peters et al. (2018), Deriu et al. (2020), Olah (2015), Miller et al. (2017), Roller et al. (2017).

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Esta línea de trabajo hace una propuesta de investigación enfocada en los desafíos relacionados con la gobernanza inteligente. Se hace desde la perspectiva de construir nuevos prototipos basados en sistemas inteligentes, mejorados

con metodología y tecnologías diferentes con el fin de exhibir "inteligencia holística". Se considerará el hardware, software, matemáticas y el conocimiento de dominio de un modo metodológico.

La hipótesis de trabajo es que se puede resolver varios problemas diferentes analizando sus características cuantitativas y cualitativas subyacentes, así como proporcionando algoritmos avanzados que pueden buscar, optimizar y aprender por sí mismos en aquellas situaciones donde el conocimiento del problema es muy limitado. Además, facilitando metodologías y procesos en particular a la gobernanza de las ciudades inteligentes y sostenibles.

Los desafíos son muchos, así que se extenderán las técnicas del estado del arte en optimización, aprendizaje máquina y análisis de datos para proporcionar aplicaciones usables y una base perdurable de algoritmos, sus fundamentos y nuevas formas de adaptarlos a las plataformas de la ciudad.

Particularmente, este trabajo se enfoca en en la construcción de un data set en áreas de Gobernanza y I&D – y el diseño y construcción de un prototipo de Chatbot aplicado a gobernanza local y un laboratorio de I&D.

3. RESULTADOS OBTENIDOS/ESPERADOS

La hipótesis de este proyecto es que aplicando Rasa NLU podemos aportar dos de las principales componentes de un flujo de diálogo en la interpretación de la pregunta y completado de respuestas, pudiendo buscar, optimizar y aprender por sí mismos en aquellas situaciones donde el conocimiento del problema es muy limitado.

Además, los resultados esperados son construir:

- Data set en áreas de Gobernanza y I&D

- Generación un informe de Chatbots inteligentes usando RASA.

- Diseño y prototipo de Chatbot aplicados a Gobernanza Local (en áreas de Empleo y de Medioambiente) y un Laboratorio de I&D (Lab-TEm) tanto para servicio de atención al cliente como descubrimiento de información en Lab-TEm.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo se encuentra formado por cuatro investigadores con distintos niveles de posgrado, dos Doctores en Ciencias de la Computación, un Magister en Ciencias de la Computación, y un estudiante de la Carrera Ingeniería en Sistemas.

Esta línea de investigación proporcionará un marco propicio para la iniciación y/o finalización de estudios de posgrado de los integrantes docentes. De igual forma, será un ámbito adecuado para la realización de tesis de grado.

BIBLIOGRAFÍA

- ADIWARDANA, D., LUONG, M., SO, D., HALL, J., FIEDEL, N., THOPPILAN, R., YANG, Z., KULSHESHTA, A., NEMADE, G., LU, Y., & LE, Q.V. (2020). Towards a Human-like Open-Domain Chatbot. ArXiv, abs/2001.09977.
- BUNK, T. VARSHNEVA, D. VLASOV, V. NICHOL, A. (2020). *DIET: Lightweight Language Understanding for Dialogue Systems*.
- DERIU, J., TUGGENER, D., DANIKEN, P.V., CAMPOS, J.A., RODRIGO, Á., BELKACEM, T., SORORA, A., AGIRRE, E., & CIELIEBAK, M. (2020). Spot The Bot: A Robust and Efficient Framework for the Evaluation of Conversational Dialogue Systems. ArXiv, abs/2010.02140.
- COURTAUDON, A. (2012). *La gestion intelligente avec le numérique : une nouvelle dynamique pour les villes et territoires français*. Blog Administration Numérique – MARKESS International.
- COLLOBERT, R. WESTON M., BOTTOU, L. KARLEN M, KAVUKCUOGLU K., and KUKSA P. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12(Aug):2493–2537, 2011.
- INSTITUTO NACIONAL DE ESTADISTICAS Y CENSO (INDEC). (2020). Acceso y uso de tecnologías de la información y la comunicación. EPH Cuarto trimestre de 2020. *Informes Técnicos Ciencia y Tecnología*, 5(1), 1-16.
- MILLER, A. H., “PARLAI: A Dialog Research Software Platform”, *ParLAI: A Dialog Research Software Platform*. arXiv e-prints, 2017.
- OLAH C. (2015). *Understanding LSTM Networks*.
- ORACLE. (2022). *What is a Chatbot?*. Oracle Cloud Infrastructure. <https://www.oracle.com/chatbots/what-is-a-chatbot/>
- PETERS M.E., Mark Neumann, IVVER M., GARDNER M., CLARK ,C. LEE K., and ZETTERMOVER L. (2018) Deep contextualized word representations. In *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2018.
- ROLLER, S., DINAN, E., GOYAL, N., Ju, D., WILLIAMSON, M., LIU, Y., Xu, J., Ott,

M., SHUSTER, K., SMITH, E.M., BOUREAU, Y., & WESTON, J. (2020). Recipes for building an open-domain chatbot. ArXiv, abs/2004.13637.

TURING A. M. (1950). Computing Machinery and Intelligence, *Mind*, 49, (433-460).

YAN Li, MANOJ A THOMAS & DAPENG LIU (2020) From semantics to pragmatics: where IS can lead in Natural Language Processing (NLP) research, European Journal of Information Systems, ELIZA: a very basic Rogerian psychotherapist chatbot.

Detección Anticipada de Riesgos en la Web

Leticia Cagnina*, M. Paula Villegas**, M. José Garciarena***, Martín Loyola****
Sergio Burdisso, Darío Funez, Horacio Thomson, Marcelo Errecalde*****

Proyecto: “Aprendizaje automático y toma de decisiones
en sistemas inteligentes para la Web” (PROICO 03-0620)

Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)

Departamento de Informática, Universidad Nacional de San Luis

Ejército de los Andes 950 - (D5700HHW) San Luis - Argentina

e-mail de contacto: {merreca}@unsl.edu.ar

Resumen

Este artículo describe, brevemente, las tareas de investigación que nuestro grupo está llevando a cabo en el área de *Detección Anticipada de Riesgos (DAR) en la Web*. Esta línea de investigación comenzó en el año 2017 con la participación de nuestro grupo en la tarea *eRisk 2017: Pilot Task on Early Detection of Depression* donde se obtuvo el mejor desempeño (de acuerdo a la medida ERDE50) sobre un total de 30 contribuciones de 8 instituciones diferentes de Francia, Alemania, USA, México, Argentina, Canadá y Rusia. A partir de ese evento, se continuó participando en forma ininterrumpida en este evento en otras tareas de DAR vinculadas a depresión, anorexia, y auto-lesiones con distintos enfoques surgidos de los trabajos de postgrado de 5 tesis de Maestría y Doctorado. En todas las participaciones del grupo, se han presentado propuestas que consituyen en la actualidad el estado del arte del área con más de 12 publicaciones en el tema.

Palabras claves: Minería de Textos, Procesamiento del Lenguaje Natural, De-

tección Anticipada de Riesgos, Sistemas Inteligentes para la Web

Contexto

La *Detección Anticipada de Riesgos en la Web* se está abordando dentro de la línea de investigación “Minería de Textos y de la Web” en el marco del Proyecto de Investigación Consolidado titulado “Aprendizaje automático y toma de decisiones en sistemas inteligentes para la Web” (PROICO 03-0620). El proyecto, aprobado por evaluadores externos a la UNSL, se desarrolla en el *Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)* de la UNSL y ha sido financiado en forma directa por la UNSL y en forma indirecta por el Programa de Incentivos (22/F637), el CONICET, a través de un investigador adjunto y 4 becas de doctorado y una beca de postgrado de la UNSL.

1. Introducción

La Detección Anticipada de Riesgos (DAR) puede considerarse como un *problema multi-objetivo* en el que el desafío es encontrar un balance adecuado entre dos aspectos diferentes y relacionados: 1) la *precisión* en la iden-

* Investigadora - CONICET

** Becaria CONICET - LIDIC

*** Becaria UNSL

**** Becario CONICET - IMASL

***** Becario CONICET - LIDIC

tificación de usuarios de riesgo y, 2) el *tiempo mínimo* que requiere la detección de un usuario de riesgo para ser *confiable*. El primer aspecto es generalmente abordado como un problema de clasificación típico y evaluado con métricas de clasificación estándar como *precisión* (en inglés *precision*), *alcance/cobertura* (en inglés *recall*) y F_1 . El segundo implica una política para decidir *cuándo* la información de un usuario catalogado como de riesgo es *suficiente* para dar la alarma/alerta y suele ser evaluado penalizando al retraso en tomar esa decisión. De hecho, las métricas de evaluación temporal utilizadas en DAR como $ERDE_\theta$ y $F_{latency}$ combinan ambos aspectos de diferentes maneras.

La DAR ha sido abordada en el contexto del Laboratorio de Predicción Temprana de Riesgos en Internet (eRisk) el cual se ocupa de la exploración de nuevos modelos de detección temprana de riesgos y metodologías de evaluación con impacto directo en aspectos sociales y de la salud [8]. El laboratorio comenzó en 2017 abordando el problema de la detección temprana de depresión en usuarios de un foro en línea (Reddit) [9]. En 2018, la detección temprana de signos de anorexia se agregó como un nuevo desafío para el laboratorio, junto con una versión ampliada de la tarea del año anterior [10]. Los datos de prueba se organizaron en 10 fragmentos (*chunks*) y fueron proporcionados a cada equipo fragmento a fragmento. Los modelos de los participantes se evaluaron utilizando la métrica de evaluación ERDE introducida por Losada y otros [8] para considerar tanto la corrección de la clasificación como la demora incurrida por el sistema para tomar la decisión. En 2019, la tarea de detección temprana de depresión fue reemplazada por dos nuevos desafíos: la detección temprana de signos de autolesiones (en inglés *self-harm*) y medir la gravedad de los signos de depresión [11]. Para esa edición del laboratorio, se consideraron nuevas medidas de desempeño. Primero, la medida de rendimiento $F_{latency}$ propuesta por Sadeque y otros [15] se incorporó como medida complementaria al ERDE. Por otro lado, se agregaron métricas de evaluación ba-

sadas en *ranking* para ayudar a los profesionales a tomar sus decisiones en problemas de la vida real. Ese año también marcó el final del procesamiento de datos basado en fragmentos. A partir de ese año, se utilizó un enfoque publicación por publicación (*post-by-post*) para las distintas tareas, que se asemeja a un escenario de la vida real donde los usuarios escriben sus mensajes de a uno por vez. En 2020 se eliminó la tarea de detección temprana de signos de anorexia pero las demás tareas se mantuvieron [12]. Finalmente, en 2021, se introdujo la tarea de detección temprana de signos de juego patológico mostrándose a continuación una breve descripción de las dos tareas en las que participó nuestro grupo de investigación:

- **Tarea 1:** Detección anticipada de signos de juego patológico. Para esta tarea, el objetivo era detectar, tan pronto como sea posible, a los usuarios que fueran jugadores compulsivos o que tuvieran patrones tempranos de juego patológico. Los datos de la tarea consistían en una serie de escritos de usuarios de medios sociales recopilados en orden cronológico. No se proporcionaron datos de entrenamiento, por lo que cada equipo tuvo que construir su propio corpus para entrenar sus modelos.
- **Tarea 2:** Detección temprana de signos de autolesión. Para esta tarea, el objetivo era el mismo que con las ediciones de eRisk 2019 y 2020, es decir, procesar secuencialmente las evidencias y detectar rastros tempranos de autolesión tan pronto como sea posible. Ese año, los datos de entrenamiento fueron la combinación de los datos de entrenamiento y prueba de la edición 2020.

El desempeño en ambas tareas se evaluó utilizando medidas de clasificación estándar (precisión, cobertura, y F_1), medidas que penalizan el retraso en la respuesta (ERDE y $F_{latency}$), y métricas de evaluación basadas en rankings. Los valores de F_1 y $F_{latency}$ se calcularon con respecto a la clase positiva. Para

calcular estas medidas, para cada publicación de cada usuario, a los modelos participantes se les pidió que proporcionaran una decisión, que indicaba si el usuario estaba en riesgo (indicado con un uno) o no (indicado con un cero), y una puntuación, que representaba el nivel de riesgo del usuario (estimado de la evidencia vista hasta ese momento). En ese contexto, si un usuario fue clasificado como de riesgo, las decisiones posteriores no fueron consideradas.

2. Líneas de Investigación y Desarrollo

Nuestros trabajos en la DAR, se relacionan con trabajos previos que abordaban el problema de procesar datos en forma secuencial y clasificarlos lo antes posible [14]. En el área específica de DAR, los principales enfoques utilizados fueron:

- Enfoque *Temporal Variation of Terms* (TVT) [5]
- Enfoque *Flexible Temporal Variation of Terms* (FTVT) [6]
- Enfoque *Sequential-Incremental Classification* (SIC) [6]
- Enfoque *k-TVT* [4]
- Enfoque *SS3*[1, 2, 3]
- Enfoque *EarlyModel*[13]
- Enfoque *EARLIEST*[13, 7]

2.1. Resultados Obtenidos

En 2017, en el marco de la Conference and Labs of the Evaluation Forum (CLEF 2017), nuestro grupo participa en el eRisk 2017 (<https://early.irlab.org/2017/index.html>) con un método diseñado específicamente para este tipo de tarea (*TVT*) obteniendo el *mejor valor* en la medida $ERDE_{50}$ [16, 5]. En 2018, se utilizan los enfoques *FTVT* y *SIC* obteniéndose los mejores valores de $ERDE_5$

tanto en *anorexia* como *depresión* y el valor más alto de precisión en *anorexia* **0.91** [6]. En los años 2019 y 2020 se participó en ambos casos con el enfoque *SS3* [1, 2, 3] obteniéndose resultados del estado del arte en el área. En 2021 [13], y a diferencia de las participaciones anteriores en los eRisk Labs, el énfasis se puso en las políticas de alerta temprana que deciden si un usuario catalogado como de riesgo debe ser efectivamente reportado como tal. Allí, propusimos tres políticas diferentes de alerta temprana para la detección temprana del riesgo de juego patológico y detección temprana del riesgo de autolesiones. El primer enfoque utiliza modelos de clasificación estándar para identificar a los usuarios de riesgo y una política de alerta temprana simple (manual) basada en reglas. El segundo enfoque es un modelo de aprendizaje profundo entrenado de extremo a extremo (end-to-end) que aprende simultáneamente a identificar a los usuarios de riesgo y la política de alerta temprana a través de un enfoque de aprendizaje por refuerzo. Finalmente, el último enfoque consiste en un modelo simple e interpretable que identifica a los usuarios de riesgo, integrado con una política global de alerta temprana. Esa política, basada en el nivel de riesgo estimado (global) para todos los usuarios procesados, decide qué usuarios deben informarse como riesgosos. Con respecto a los resultados alcanzados, nuestros modelos obtuvieron el mejor rendimiento en términos de métricas de rendimiento basadas en decisiones (F_1 , $ERDE_{50}$, $F_{latency}$) así como en términos de medidas de rendimiento basadas en ranking, para ambas tareas. Además, en términos de la medida de $F_{latency}$, el rendimiento obtenido en la primera tarea fue el doble que el segundo mejor equipo.

3. Formación de Recursos Humanos

Trabajos de tesis vinculados con las temáticas descritas previamente:

- 1 tesis de Maestría en ejecución con beca de postgrado de la UNSL.
- 3 tesis de Doctorado en ejecución con becas de CONICET.
- 1 tesis de Doctorado finalizada con beca de CONICET.

Referencias

- [1] S. G. Burdisso, M. Errecalde, and M. Montes-y Gómez. A text classification framework for simple and effective early depression detection over social media streams. *Expert Systems with Applications*, 133:182 – 197, 2019.
- [2] S. G. Burdisso, M. Errecalde, and M. Montes-y Gómez. UNSL at eRisk 2019: a unified approach for anorexia, self-harm and depression detection in social media. In *Working Notes of CLEF 2019*, Lugano, Switzerland, 2019. CEUR Workshop Proceedings.
- [3] S. G. Burdisso, M. Errecalde, and M. Montes-y Gómez. τ -SS3: A text classifier with dynamic n-grams for early risk detection over text streams. *Pattern Recognition Letters*, 138:130 – 137, 2020.
- [4] L. C. Cagnina, M. L. Errecalde, M. J. Garciarena Ucelay, D. G. Funez, and M. P. Villegas. *k*-tvt: a flexible and effective method for early depression detection. In *XXV Congreso Argentino de Ciencias de la Computación. CA-CIC 2019. Libro de actas*, pages 547–556, 2019.
- [5] M. L. Errecalde, M. P. Villegas, D. G. Funez, M. J. Garciarena Ucelay, and L. C. Cagnina. Temporal variation of terms as concept space for early risk prediction. In *Working Notes of the Conference and Labs of the Evaluation Forum - CEUR Workshop Proceedings*, volume 1866, 2017.
- [6] D. G. Funez, M. J. Garciarena Ucelay, M. P. Villegas, S. G. Burdisso, L. C. Cagnina, M. Montes y Gomez, and M. L. Errecalde. Unsl’s participation at erisk 2018 lab. In *Working Notes of the Conference and Labs of the Evaluation Forum - CEUR Workshop Proceedings*, volume 2125, 2018.
- [7] T. Hartvigsen, C. Sen, X. Kong, and E. Rundensteiner. Adaptive-halting policy network for early classification. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 101–110, 2019.
- [8] D. E. Losada and F. Crestani. A test collection for research on depression and language use. In *Proc. of Conference and Labs of the Evaluation Forum (CLEF 2016)*, pages 28–39, Evora, Portugal, 2016.
- [9] D. E. Losada, F. Crestani, and J. Parapar. erisk 2017: Clef lab on early risk prediction on the internet: experimental foundations. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 346–360. Springer, 2017.
- [10] D. E. Losada, F. Crestani, and J. Parapar. Overview of erisk: early risk prediction on the internet. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 343–361. Springer, 2018.
- [11] D. E. Losada, F. Crestani, and J. Parapar. Overview of erisk 2019 early risk prediction on the internet. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 340–357. Springer, 2019.
- [12] D. E. Losada, F. Crestani, and J. Parapar. Overview of erisk at clef 2020: Early risk prediction on the internet (extended overview). 2020.

- [13] J. M. Loyola, S. G. Burdisso, H. Thompson, and M. L. Errecalde. Unsl at erisk2021: A comparison of three early alert policies for early risk detection. In *Working Notes of the Conference and Labs of the Evaluation Forum - CEUR Workshop Proceedings*, volume 2936, 2021.
- [14] J. M. Loyola, M. L. Errecalde, H. J. Escalante, and M. M. Gomez. Learning when to classify for early text classification. In *Argentine Congress of Computer Science*, pages 24–34. Springer, 2017.
- [15] F. Sadeque, D. Xu, and S. Bethard. Measuring the latency of depression detection in social media. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 495–503, 2018.
- [16] M. P. Villegas, D. G. Funez, M. J. Garcianena Ucelay, L. C. Cagnina, and M. L. Errecalde. Lidic - unsl’s participation at erisk 2017: Pilot task on early detection of depression. In *Working Notes of CLEF 2017 - CEUR Workshop Proceedings*, volume 1866, 2017.

ANÁLISIS DE CALIDAD DE ARENAS DE FRACTURACIÓN MEDIANTE VISIÓN ARTIFICIAL Y REDES NEURONALES

Mg. Ing. Carlos Gustavo Rodriguez Medina ¹, Dr. Ing. Oscar Daniel Chuk ²,
Lic. Adriana Luna, Ing. Regina Bertero, Ing. Enrique Núñez.

Instituto de Investigaciones Mineras / Facultad de Ingeniería /
Universidad Nacional de San Juan

Av. Libertador Gral. San Martin 1109 oeste. San Juan
0264-4211700 (int. 285 ¹, int. 389 ^{1,2})

grodriguez@unsj.edu.ar ¹, dchuk@unsj.edu.ar ²

RESUMEN

La fracturación hidráulica realizada para extraer hidrocarburos de yacimientos no convencionales requiere de la inyección de arenas que actúan como apuntalantes de la fractura. La calidad de las arenas para cumplir dicha función se evalúa por la norma API19C, que establece medidas geométricas – esfericidad y redondez- más un porcentaje de fractura cuando la muestra es sometida a compresión. Todas estas mediciones deben encontrarse dentro de valores límites para asegurar la calidad de la arena.

Según la norma, la inspección se realiza visualmente por un operador sobre un conjunto de 20 granos. Esto introduce un importante grado de subjetividad, y poca validez estadística. Para solucionar esto, la bibliografía refiere distintos métodos basados en visión artificial. Cada uno de estos métodos tiene ventajas y desventajas según la geometría, color de la partícula y la definición de la imagen usada.

En este trabajo se presenta una metodología integrada a partir de varios métodos conocidos, más uno novedoso desarrollado por los autores para medir la redondez, que es la variable más difícil de medir. Las distintas medidas son tratadas por redes neuronales para dar una medida final de la redondez, que tiene un alto grado de correlación con la

medida teórica de cada partícula considerada. Las medidas de esfericidad y porcentaje de fractura obtenidas también han dado valores consistentes.

El método de visión artificial desarrollado es sumamente eficiente para determinar, a partir de las medidas obtenidas, la capacidad de una arena para actuar como apuntalante en una operación de fractura hidráulica.

Palabras clave: Visión artificial, Redes neuronales, Fracturación hidráulica, Arenas, geometría.

CONTEXTO

Este trabajo presenta de manera resumida parte de los avances y resultados del Proyecto de Desarrollo Tecnológico y Social (PDTs – UNSJ) “*Desarrollo de un algoritmo de análisis de calidad de arenas de fracturación basado en visión artificial*”, como así también del Proyecto de Investigación Científica (PIC – UNSJ) “*Clasificación de partículas de arena de fracturación mediante redes neuronales convolucionales*”, ambos en ejecución durante el periodo 2020 – 2022.

Tales proyectos se inscriben dentro de la línea de investigación que lleva adelante un grupo de trabajo en la temática de Inteligencia y Visión Artificial orientado al control automático de procesos mineros, en el

Instituto de Investigaciones Mineras de la Facultad de Ingeniería de la UNSJ.

1. INTRODUCCIÓN

Para que las partículas de arena utilizadas en fracturación hidráulica (fracking) actúen apropiadamente como apuntalante o “propante”, deben resistir las altas presiones usadas en el proceso, del orden de los 10.000 psi. Esta capacidad es clave para el éxito del proceso. Para determinar la resistencia de la arena, se reconocen como base las normas del American Petroleum Institute API RP 19 B, la cual define un ensayo a la compresión para determinar el porcentaje de fractura. Pero también son de rigor, según la norma API 19C [1], el análisis de distribución granulométrica y, fundamentalmente, las medidas geométricas de redondez y esfericidad. Éstas últimas son definitorias pues si no superan un valor de 0.6 (ambas medidas oscilan entre 0 y 1), la arena es descartada.

Según las normas citadas, la determinación de la redondez y esfericidad recae en la observación de 20 granos tomados al azar por parte de un operador, por asimilación a las imágenes de una cartilla de referencia propuesta por Krumbein y Sloss [2]. Estos autores desarrollaron sus medidas en base a las expresiones de Wadell [3] [4] [5], pero la aplicación de sus fórmulas es muy complicada en la práctica; de allí la generación de la cartilla y la intervención de un observador humano. Esta metodología padece de una fuerte dependencia del criterio de dicho observador, lo cual deriva en resultados sustancialmente disímiles que invalidan el procedimiento, aunque es el que se sigue por norma.

El uso de la visión artificial es una alternativa poderosa para lograr una medida objetiva de las propiedades geométricas de partículas, de lo cual Rodríguez [6] presenta una detallada revisión del estado del arte.

La bibliografía existente cubre la determinación de la distribución granulométrica [7], como así también esfericidad [8] y la redondez [9]-[10], sin

dejar de lado otros indicadores como el factor de forma, la concavidad y la convexidad [11]. Entre todas las medidas detalladas, la redondez es la que es más susceptible de error por la apreciación humana. Esto ha llevado por un lado a intentar medir esta propiedad por técnicas estrictamente geométricas. Rodríguez et al. [13] registran 19 formas distintas de medir la redondez con esta perspectiva.

Por otro lado, se han aplicado técnicas indirectas para determinar la redondez usando como datos de entrenamiento de un algoritmo estimador una serie de mediciones previamente hechas con algún criterio de los citados en el párrafo anterior. Sun [14], por ejemplo, utiliza un algoritmo de optimización de enjambre de partículas para calcular la redondez.

Se presenta en este artículo un abordaje combinado para la determinación de la redondez. Dado que los distintos métodos adolecen de ciertos problemas en determinadas circunstancias, las mediciones logradas por métodos tradicionales más dos desarrolladas por los autores se combinan en entradas de distintas redes neuronales para rescatar lo mejor de cada procedimiento y dar un resultado final confiable.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

En el Laboratorio de Control Automático del Instituto de Investigaciones Mineras (Facultad de Ingeniería de la Universidad Nacional de San Juan) se realiza investigación y desarrollo respecto a la automatización de procesos de molienda de minerales aproximadamente desde el año 2000. Desde hace algunos años se viene desarrollando el Área de Procesamiento de Imágenes.

Actualmente se está trabajando en la caracterización de arenas especiales (calidad, forma, color, tamaños, clasificación del tipo de material, etc.) mediante el procesamiento de imágenes para ser utilizadas en la extracción de petróleo mediante el método de *Fracking*.

3. RESULTADOS OBTENIDOS

En este estudio se usará como medida de referencia de la redondez la establecida por Wadell [3]:

$$R_w = \frac{\sum_{i=1}^N r_i / N}{r_{\max_in}} \quad (1)$$

Donde r_i es el radio de curvatura de las esquinas, definiéndose las mismas como “aquellas partes del contorno de un área que tiene un radio de curvatura menor o igual que el radio de curvatura del máximo círculo inscripto r_{\max_in} en la misma área”.

El cálculo manual de esta fórmula es tedioso, por lo cual se recurre a métodos alternativos como el visual establecido por la norma API 19C y los que se presentarán a continuación basados en visión artificial. Aun así, el cálculo manual a partir de (1) se ha realizado en este estudio en 572 figuras que se han usado como referencia, a fin de poder entrenar redes neuronales que se alimentan con datos alternativos, los cuales se describirán a continuación. Estas figuras cubren todo el rango habitual de mediciones de redondez entre 0.1 y 1, y están combinadas con esfericidad que también cubren todo el rango posible.

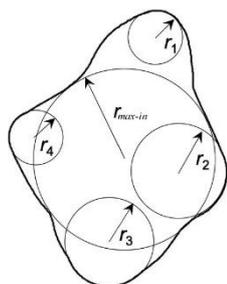


Figura 1: Definición de radios de curvaturas de esquinas y máximo círculo inscripto.

Si bien algunos de los procedimientos que se han usado por visión artificial – y que se describirán a continuación– recurren estrictamente a la expresión (1), ciertas dificultades particulares de los mismos hacen que sea necesario recurrir a medidas alternativas que complementen las mismas a fin de entregar un valor final representativo. Todos estos datos son ingresados a distintas

redes neuronales que entregarán finalmente el valor correcto de R_w .

Los datos básicos obtenidos por visión artificial, usados para alimentar las redes neuronales son los siguientes:

1) Perímetro de la elipse equivalente $pEeq$. Es la elipse que minimiza la diferencia de superficie entre la misma y la partícula en consideración. Es usada en este estudio como una medida de la definición de la imagen de la partícula. Si $pEeq$ es suficientemente grande, se puede asegurar que el contorno de la partícula contendrá la suficiente cantidad de píxeles como para realizar cálculos por la expresión (1); caso contrario habrá que recurrir a medidas complementarias.

2) Redondez basada en el perímetro R_p . Está definida como [12]

$$R_p = (pEeq / P)^2 \quad (2)$$

Donde P es el perímetro en píxeles de la partícula analizada. Si la misma fuera perfectamente elíptica, entonces $pEeq = P$ y en consecuencia $R_p = 1$. En cambio, a medida que la partícula presenta convexidades y concavidades, $pEeq < P$ y por lo tanto $R_p < 1$. Esta es una medida de redondez que no se corresponde con la Wadell R_w (1), pero es muy consistente para bajas redondeces y por lo tanto útil para alimentar una red neuronal que luego entregará el valor de R_w .

3) Relación entre el perímetro del máximo círculo inscripto al perímetro de la figura. Definida como

$$P_{mci}2P = 2 \pi r_{\max_in} / P \quad (3)$$

Esta medida, aunque no definitiva, es complementaria e indicativa de la rugosidad de la partícula, al igual que (2), y por lo tanto útil para complementar la información que entra a las redes neuronales. Mientras más oscilaciones tenga el contorno (menor redondez), mayor será P y menor será $P_{mci}2P$.

4) Esfericidad de ejes E_e

$$E_e = mal / Mal \quad (4)$$

Donde a y b son los ejes mayor y menor de la elipse equivalente de la figura considerada. Esta medida también es introducida como información complementaria, debido a que es sabido que las medidas alternativas como (2) sufren alteraciones debidas al nivel de esfericidad.

5) Cálculo de redondez según Wadell (1) R_{wz} usando el algoritmo de Zheng y Hryciw [15]. Este algoritmo detecta las zonas de convexidad de la partícula por técnicas estrictamente geométricas, básicamente trazando una línea entre distintos puntos del contorno y detectando si dicha línea queda incluida en la partícula o fuera de ella. Luego, un conjunto de puntos convexos cercanos entre sí (Figura 2b) son tomados para encontrar por mínimos cuadrados el círculo que mejor se ajusta a dicho puntos [16].

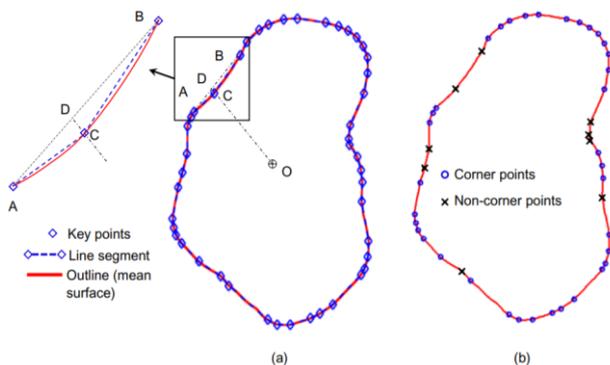


Figura 2: Identificación de esquinas según el método de Zheng y Hryciw.

La desventaja de este método es que la detección de zonas convexas puede fallar si la imagen de la partícula no tiene suficiente definición.

6) Cálculo de redondez según Wadell (1) R_{wc} detectando las zonas convexas por medio de la descripción polar de la partícula (propuesta de los autores). Véase la Figura (3).

En este caso la detección de las esquinas se realiza mediante el análisis de la derivada segunda de $\rho(\varphi)$, y una vez detectados los puntos correspondientes se realiza un ajuste de círculos en una forma semejante al caso anterior. Los resultados son semejantes al algoritmo de Zheng, pero mientras aquel tiene el problema ya descrito en relación con la definición en píxeles de la imagen, este procedimiento tiene el problema de que no se

puede aplicar cuando hay reentrancia del radio polar, es decir cuando hay más de un radio ρ para un determinado ángulo φ , o sea, cuando $\rho(\varphi)$ no es biunívoca.

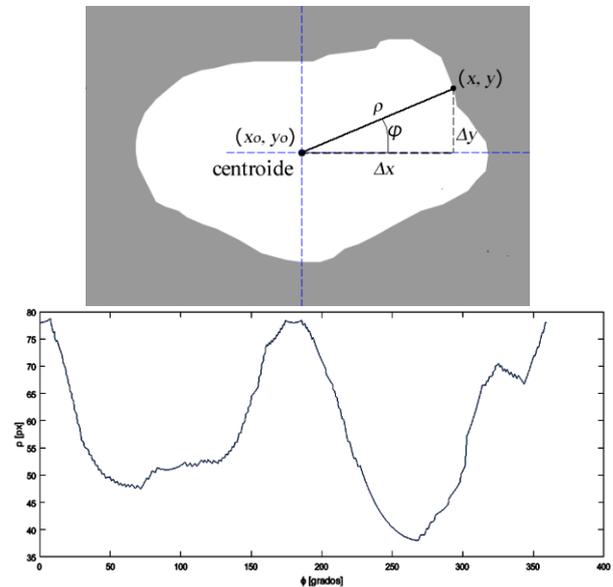


Figura 3: Descripción polar del contorno

Los procedimientos 5) y 6), entonces, aunque en muy pocos casos, pero pueden fallar, y en dichos casos hay que alimentar las redes neuronales con los datos 1) a 4), como se muestra a continuación.

Definición de las redes neuronales

A los fines de llegar a un cálculo final, se definen cuatro redes neuronales según la tabla siguiente.

	Net1	Net 2	Net 3	Net 4
Condiciones	Existen R_{wc} y R_{wz}	No existe R_{wc}	No existe R_{wz}	No existe ni R_{wc} ni R_{wz}
Entradas	1. pEq	1. pEq	1. pEq	1. pEq
	2. Rp	2. Rp	2. Rp	2. Rp
	3. R_{wc}	3. R_{wz}	3. R_{wc}	3. P_{mci2P}
	4. R_{wz}	4. P_{mci2P}	4. P_{mci2P}	4. Ee
	5. P_{mci2P}	5. Ee	5. Ee	
	6. Ee			

Tabla 1: Definición de entradas de las redes neuronales de cálculo

Las cuatro redes neuronales han sido entrenadas utilizando como datos de

entrenamiento las redondeces calculadas manualmente de las 572 figuras de referencia citadas previamente.

Dado que en la gran mayoría de los casos el cálculo de R_{wc} y R_{wz} es posible, la red más usada para el cálculo final es Net1.

La correlación obtenida en la medición de las figuras de referencia se muestra en la Figura 4, donde se observa un buen ajuste de las mediciones.

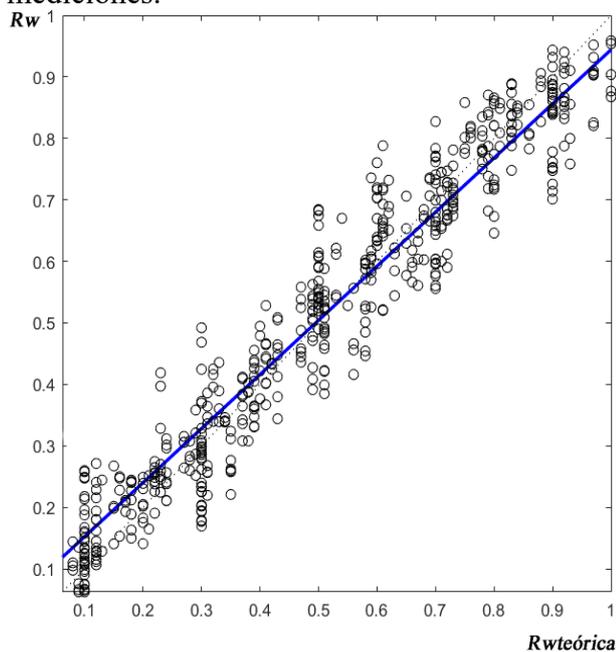


Figura 4: Ajuste de los datos obtenidos a la medida teórica.

El Coeficiente de correlación obtenido es de 0.947, y la recta de ajuste es

$$R_w = 0.88 R_{wteórica} + 0.064 \quad (5)$$

Basados en estos valores, se puede afirmar que el procedimiento es confiable.

Al momento de realizar el análisis de una muestra, el procedimiento por visión artificial permite considerar muchas más partículas (típicamente entre 200 y 500) en lugar de las 20 establecidas por la norma API. Esto le da una validez estadística de gran confiabilidad, y las diferencias de los promedios obtenidos por el procedimiento descrito en este artículo y los teóricos encontrados manualmente oscilan entre el 1% y 5%.

Otro tanto se puede decir de las otras variables geométricas necesarias para definir la calidad de la arena, a saber, la esfericidad y la distribución granulométrica.

Se ha logrado entonces un procedimiento de medición seguro, repetitivo, rápido y que no depende de la subjetividad de un operador.

Si bien es preferible que las imágenes necesarias para alimentar el algoritmo provengan de un microscopio con alto aumento, acompañado de una cámara de alta precisión, del orden de los 5Mpíxeles, los resultados obtenidos con imágenes de poca precisión –un microscopio USB estándar de bajo costo- han dado resultados también con el mismo nivel de error promedio final.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de investigación está conformado por Ingenieros Electrónicos, Electromecánicos, en Minas, Metalurgistas y Geólogos, con formación diversa en posgrados, tal como Doctorado en sistemas de control, Maestría en informática, Especialización en gestión y vinculación tecnológica, etc.

Actualmente, dos integrantes del equipo de trabajo se encuentran desarrollando su Tesis de Doctorado en Ingeniería de Procesamiento de Minerales, siendo su director de Tesis otro de los integrantes del grupo de investigación.

5. BIBLIOGRAFIA

- [1] ANSI/API (2008). Recommended practice 19C -First Edition, - Part 2: Measurement of properties of proppants used in hydraulic fracturing and gravelpacking operations.
- [2] Krumbein, W.C.; Sloss, L.L. (1963). Stratigraphy and Sedimentation; 2nd ed. W.H. Freeman: San Francisco.
- [3] Wadell, H. (1932). Volume, Shape, and Roundness of Rock Particles. Jour. Geol., XL (5), 443–451.
- [4] Wadell, H. (1933). Sphericity and Roundness of Rock Particles. Jour. Geol., XLI(3), 310–331.

- [5] Wadell, H. (1935). Volume, Shape and Roundness of Quartz Particles. *Jour. Geol.* XLIII(3), 258–280.
- [6] Rodriguez, J.M. (2013). Importance of the Particle Shape on Mechanical Properties of Soil Materials. *Licenciate Thesis*, Luleå University of Technology, Luleå.
- [7] Andersson, T. (2010). Estimating Particle Size Distributions Based on Machine Vision. *Doctoral Thesis*, Luleå University of Technology, Luleå.
- [8] Mora, C.K. (2000) Sphericity, Shape Factor, and Convexity Measurement of Coarse Aggregate for Concrete Using Digital Image Processing. *Cement and Concrete Research* 30(3), 351–358.
- [9] Roussillon, T.; Piégay, H.; Sivignon, I.; Tougne, L.; Lavigne, F. (2009). Automatic Computation of Pebble Roundness Using Digital Imagery and Discrete Geometry 35, 1992–2000, doi:10.1016/j.cageo.2009.01.013.
- [10] Drevin, G.R. (2007). Computational Methods for the Determination of Roundness of Sedimentary Particles. *Math Geol* 38, 871–890, doi:10.1007/s11004-006-9051-y.
- [11] Altuhafi, F.; O’Sullivan, C.; Cavarretta, I. (2013). Analysis of an Image-Based Method to Quantify the Size and Shape of Sand Particles. *J. Geotech. Geoenviron. Eng.* 139, 1290–1307, doi:10.1061/(ASCE)GT.1943-5606.0000855.
- [12] Charalambous, C. (2015). On the Evolution of Particle Fragmentation With Applications to Planetary Surfaces. *Doctoral Thesis*, Imperial College London.
- [13] Rodriguez, J.; Edeskär, T.; Knutsson, S. (2013). Particle Shape Quantities and Measurement Techniques - A Review. *Electronic Journal of Geotechnical Engineering* 18(A), 169–198.
- [14] Sun, T.H. (2009). Applying Particle Swarm Optimization Algorithm to Roundness Measurement. *Expert Systems with Applications* 36, 3428–3438, doi:10.1016/j.eswa.2008.02.072.
- [15] Zheng, J.; Hryciw, R. (2015). Traditional Soil Particle Sphericity, Roundness and Surface Roughness by Computational Geometry. *Géotechnique* 65, 494–506, doi:10.1680/geot.14.P.192.
- [16] Gander, W.; Golub, G.; Strebler, R. (1994). Least-Squares Fitting of Circles and Ellipses. *BIT* 34, 558–578, doi:10.1007/BF01934268.
- [17] Chuk, O.D.; Rodriguez Medina, C.G.; Luna, A. (2021). Roundness Calculus of Frack-ing Sands Using Artificial Vision and Fourier Transform. *Anales de XV Jornadas de Tratamiento de Minerales*; UNSJ, San Juan, Argentina.

MARCO DE TRABAJO DE RASGOS BIOMÉTRICOS EN QUEILOSCOPIA MEDIANTE EL USO DE MACHINE LEARNING

Sabelli, Agustín¹; Chatterjee, Parag¹; Pollo Cattaneo, Ma. Florencia¹

¹Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS)

Programa Maestría en Ingeniería en Sistemas de Información.

Universidad Tecnológica Nacional. Facultad Regional Buenos Aires.

Medrano 951 (C1179AAQ) Ciudad Autónoma de Argentina. Buenos Aires Tel +54 11 4867-7511

agustinsabelli@hotmail.com, parag@frba.utn.edu.ar, flo.pollo@gmail.com

Resumen

La Queiloscopía es el estudio de las impresiones labiales que se producen a través del análisis de las líneas, fisuras, arrugas y estrías presentes en el labio [1]. “Queilos” proviene del griego que significa labio y “scopia” examinar [2]. Según Cardoso [3], fue el antropólogo R. Fischer el pionero en esta área. Éste describió los surcos en 1902, pero no fue hasta 1932 que Edmond Locard, reconocido criminalista francés, recomendó su uso para la identificación. No obstante, tuvieron que pasar veintiocho años para que en 1950 LeMonyne Snyder los utilice en un caso real.

Aunque la Queiloscopía es un campo relativamente nuevo entre la gran cantidad de herramientas de identificación disponible para expertos forenses, de ésta se obtiene información sumamente útil como la identidad de una persona. Esto se debe a que permanecen relativamente estables y muestran diferencias en cuanto al género [4].

La Queiloscopía es un procedimiento manual donde se utilizan herramientas como lupas y escalas para analizar las huellas labiales. Esto lo convierte en una metodología propensa a errores humanos [5]. Para evitar esto y automatizarla, se precisa de un algoritmo.

Por otro lado, Machine Learning (ML) es un subcampo de la Inteligencia Artificial (IA). Esta última se define como la inteligencia exhibida por una entidad artificial para resolver problemas complejos. Tal sistema generalmente supone ser una computadora o máquina [6]. Dicho de otra forma, se puede decir que la IA es la habilidad que tiene dicha entidad de utilizar algoritmos para aprender de los datos y usar este conocimiento para tomar decisiones como lo haría un ser humano. A diferencia de este último, las máquinas que cuentan con IA corren con la ventaja de no precisar de descansos, analizar enormes cantidades de datos de forma simultánea y contar con una baja tasa de error [7].

Si bien la IA y ML han estado presentes desde hace mucho tiempo, es solamente ahora que se cuenta con el poder computacional para efectivamente desarrollar Redes Neuronales Artificiales (RNA) lo suficientemente poderosas en un lapso de tiempo razonable [8]. En el campo de la biometría, ML resalta por su capacidad de aumentar la precisión en el proceso de identificación. Las características biométricas tomadas en primera instancia no son siempre iguales a las tomadas una segunda vez. En consecuencia, el uso de técnicas de aprendizaje automático como redes

neuronales, lógica difusa, informática evolutiva, etc., ha incrementado su demanda.

En este contexto, el objetivo del proyecto es definir un marco de trabajo, utilizando ML, para determinar rasgos biométricos suaves de una persona, como el sexo y edad, a través de sus impresiones labiales.

Palabras clave: queiloscopía, machine learning, inteligencia artificial.

Contexto

La Universidad Tecnológica Nacional ha definido diversos Programas de Investigación, Desarrollo e Innovación, entre los cuales se encuentra el Programa de Sistemas de Información e Informática que tiene como objetivo “intensificar y focalizar las acciones tendientes a fortalecer y promover el crecimiento de temáticas de investigación en Sistemas de Información e Informática, y promover la interacción con la industria informática en general y de desarrollo de sistemas de información y de software en particular” [9].

En la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional (UTN.BA) se ha conformado en el año 2009 el Grupo GEMIS, con dependencia del Departamento de Ingeniería en Sistemas de Información, integrado por un equipo de docentes, alumnos y graduados con interés en la sistematización de conocimientos y su promoción sobre el campo de la Ingeniería en Sistemas de Información y la Ingeniería de Software incluyendo sus aplicaciones y abordajes metodológicos en todo tipo de escenarios.

Introducción

Los documentos de investigación sobre el uso de impresiones labiales en odontología forense son escasos en comparación con la literatura sobre la práctica generalizada de confiar en las huellas digitales para la identificación personal [10]. No obstante, recientemente ha habido un aumento de estudios desde el 2001 que demuestran un renovado interés en el estudio de éstas [11]. En una era criminalística en donde los métodos que emplean programas informáticos facilitan enormemente la labor del criminalista y reducen la subjetividad que gobierna las decisiones humanas, resulta de gran importancia hacer referencia a la cuestión en lo que se refiere a la Queiloscopía [3].

Por otro lado, la humectación continua de la cavidad oral y la secreción de sebo de las glándulas sebáceas conduce a la formación de huellas labiales latentes [12], las cuales no son visibles a diferencia de las huellas labiales visibles debido a la utilización de lápiz labial [13]. Éstas pueden ser recolectadas hasta 30 días luego de ser producidas [14].

Al ser un procedimiento sumamente sencillo y económico [13], la aplicación de la Queiloscopía resulta ventajosa y sumamente útil cuando se encuentran disponibles dichas impresiones labiales en objetos o pertenencias, como tazas, vasos, colillas, servilletas de papel o incluso en cojines u objetos similares utilizados en casos de asfixia. Cuando una huella labial latente está relacionada con una víctima o sospechoso en la escena del crimen, se convierte en evidencia y debe analizarse como tal [1]. Investigaciones incluso sugieren que este tipo de estudio podría cumplir un gran papel en la comparación, análisis e identificación exitosa de una persona en tal circunstancia [15]. De hecho, según Sharma et al. [12], para

asociar a alguien con algo o un lugar en específico, la Queiloscopía representa una de las formas más confiables de realizarlo.

Además, la Queiloscopía puede apoyarse de otros datos biométricos para lograr una identificación exitosa [11], como ser las huellas dactilares [16]. Por otro lado, en lo que respecta a la determinación del sexo se ha demostrado que la Queiloscopía posee un alto grado de exactitud y reproducibilidad para predecirla [12].

Según Stamm [17], una situación redundante en los países latinoamericanos es el hecho de que no está protocolizado el empleo de la Queiloscopía como sistema de identificación de personas. No hay concientización entre los profesionales odontólogos de la implementación de registros de huellas labiales, situación que ameritaría una revisión, ya que se trata de un método que reúne los requisitos de todo sistema de identificación: unicidad, perennidad, invariabilidad y clasificabilidad.

Lamentablemente, en la actualidad no existen muchas bases de datos que contengan imágenes de impresiones de labios y aquellas que están abiertas al público para la investigación son escasas [5]. Stamm [17] también resalta la importancia de fomentar la implementación de bases de datos Queiloscópicas como procedimiento de rutina en la práctica diaria del odontólogo, instando a un rediseño de la Historia Clínica Odontológica, en razón de la cantidad y, sobre todo, calidad de información que debería contener.

No existe un estándar para obtener las impresiones de labios. Muchos trabajos de investigación abogan el uso de distintas técnicas como fotografiarlas y el uso de lápices labiales u otros agentes colorantes [5]. Sumado a eso, los métodos generalmente utilizados para llevar a cabo el análisis son manuales e involucran el

uso de un software de edición de imagen como el Adobe Photoshop® [2], [11], [12], [18], [19].

A pesar de que métodos automáticos para el reconocimiento que utilizan algoritmos de Deformación Dinámica del Tiempo (DTW), Transformada Top-Hat, métodos de conteo de votos (“vote counting”) y la Transformada de Hough han resultado ser bastante eficientes [5], no se ha incursionado demasiado en la obtención de otros datos como ser el género de la persona. Cardoso [3] menciona que también un aspecto a mejorar es proporcionar un margen de error intrínseco a su propia metodología, una suerte de “likelihood ratio” (LR) que expresa la probabilidad de que el resultado sea adecuado. Cumpliendo esto último le permitiría ser utilizado en cualquier informe pericial.

Líneas de Investigación, Desarrollo e Innovación

El proyecto presentado en el presente trabajo propone avanzar en el estado de la implementación de Machine Learning para la solución de problemas de identificación de individuos.

En el período 2016-2018 se llevó adelante un Proyecto de Investigación y Desarrollo (PID) denominado “Implementación de sistemas inteligentes para la asistencia de alumnos y docentes de la carrera de Ingeniería en sistemas de información” que indaga en el uso de tecnologías de IA y del desarrollo de software inteligente para la resolución de problemas en el ámbito educativo.

Luego, en el período 2019-2021, se llevó adelante un Proyecto de Investigación y Desarrollo (PID) denominado “Prácticas ingenieriles aplicadas para la implementación de Sistemas Inteligentes basados en Machine

Learning” que indaga en el uso de buenas prácticas ingenieriles para el abordaje de desarrollo de software inteligente que utiliza ML. En esta línea se busca validar diferentes soluciones a problemas en diferentes ámbitos. He aquí la participación de la presente propuesta de trabajo.

Como una nueva línea en el marco de los proyectos desarrollados, se propone ahondar en las particularidades de tareas o áreas específicas de la industria del software y la Inteligencia Artificial, mediante el uso de Machine Learning en el marco de las buenas prácticas ingenieriles.

De esta manera, se articula dentro de los objetivos de GEMIS en el campo de la Informática, la generación de nuevos conocimientos en el área de la Ingeniería de Software y la Inteligencia Artificial.

Resultados esperados

Objetivo General

Definir un método semiautomático que, utilizando algoritmos de Machine Learning, ayude a expertos forenses a poder determinar rasgos biométricos suaves de una persona, como el sexo y edad, a través de sus impresiones labiales.

Objetivos Específicos

Los objetivos específicos son:

- Definir los pasos que se deben llevar a cabo para limpiar y pre-procesar la impresión labial digitalizada.
- Especificar cuáles son las características a ser extraídas e inyectadas a los algoritmos de ML
- Proponer, cuantificar y comparar la efectividad de diversos algoritmos de ML para estimar el sexo y edad de una persona

Resultados esperados

Como resultado de las tareas a desarrollar se espera:

- Definir un procedimiento para pre-procesar y seleccionar las características más influyentes de las impresiones que luego serán inyectadas a los algoritmos de ML encargados de determinar los rasgos biométricos suaves
- Contar con un marco de trabajo para ayudar a los expertos forenses durante una investigación criminal a mejorar el proceso de identificación de una persona a través de sus impresiones labiales;
- Vincular con otros PID del Grupo, relacionados con Machine Learning, Sistemas Expertos y Sistemas Inteligentes.

Formación de Recursos Humanos

El equipo se encuentra conformado por investigadores formados, investigadores de apoyo, graduados de grado y alumnos de la carrera de Ingeniería en Sistemas de Información. Esta nueva línea de trabajo permitirá:

- la generación nuevas oportunidades y experiencias para la formación de los investigadores;
- la incorporación de nuevos becarios graduados a través de las becas BINID;
- el desarrollo de por lo menos 2 tesis de la Maestría en Ingeniería en Sistemas de Información (1 en proceso). Además, se considera vincular por lo menos 2 trabajos

finales de la Especialidad en Ingeniería en Sistemas de Información (uno finalizado y otro en comienzo).

Referencias

[1] I. G. Cavalcanti Caputo *et al.*, “Cheiloscopy in the human identification,” *FRCIJ*, vol. 6, no. 5, pp. 371–374, Nov. 2018, doi: 10.15406/frcij.2018.06.00231.

[2] R. Gugulothu, R. Alaparathi, dr. kotyanaik maloth, S. Kesidi, V. Kundoor, and M. Palutla, “Personal identification and sex determination using cheiloscopy,” *Journal of Indian*

[3] M. C. H. Cardoso, Queiloscopía. Método de identificación del ser humano a partir de las huellas labiales. Sociedad Mexicana de Criminología capítulo Nuevo León, 2019.

[4] A. Kumar, S. N. Prasad, V. Kamal, S. Priya, M. Kumar, and A. Kumar, “Importance of Cheiloscopy,” *IJO CR*, vol. 4, pp. 48–52, 2016, doi: 10.5005/jp-journals-10051-0012.

[5] S. Sandhya and R. Fernandes, “Lip Print: An Emerging Biometrics Technology - A Review,” in 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICIC), Dec. 2017, pp. 1–5. doi: 10.1109/ICIC.2017.8524457.

[6] J. Borana, “Applications of Artificial Intelligence & Associated Technologies,” p. 4, 2016.

[7] L. Rouhiainen, *Artificial Intelligence: 101 Things You Must Know Today About Our Future*. Lasse Rouhiainen, 2018.

[8] P. Bhatt, “MACHINE LEARNING FORENSICS: A NEW BRANCH OF DIGITAL FORENSICS,” *International Journal of Advanced Research in Computer Science*, vol. 8, pp. 217–222, Aug. 2017, doi: 10.26483/ijarcs.v8i8.4613.

[9] Universidad Tecnológica Nacional (2016). Resolución Nro. 2508/16

[10] Y. Verma *et al.*, “A study of lip prints and its reliability as a forensic tool,” *Natl J Maxillofac Surg*, vol. 6, no. 1, pp. 25–30, 2015, doi: 10.4103/0975-5950.168227.

[11] W. Furnari and M. N. Janal, “Cheiloscopy: Lip Print Inter-rater Reliability,” *Journal of Forensic Sciences*, vol. 62, no. 3, pp. 782–785, 2017, doi: 10.1111/1556-4029.13308.

[12] B. Sharma, V. Gutpa, H. Vij, E. Sharma, N. Tyagi, and S. Singh, “Cheiloscopy: A tool for antemortem identification,” *Indian J Dent Sci*, vol. 9, no. 3, pp. 176–180, 2017.

[13] A. Dolly, C. Rodrigues, R. Bankur, P. A. Gopinathan, R. Sharma, and A. Doddamani, “Evaluation of Efficacy of Three Different Materials Used in Cheiloscopy –A Comparative Study,” *J Clin Diagn Res*, vol. 10, no. 10, pp. ZC67–ZC71, Oct. 2016, doi: 10.7860/JCDR/2016/21410.8653.

[14] N. Sinha, “Resin Luting Agents-A Review,” *Journal of Dental Science and Research*, Jan. 2015.

[15] L. Chatra, T. Peter, and A. Ahsan, “Cheiloscopy,” *Int J Forensic Odontol*, vol. 1, no. 2, p. 48, 2016, doi: 10.4103/2542-5013.194272.

[16] M. Abidullah, M. N. Kumar, K. D. Bhorgonde, and D. S. P. Reddy, “Cheiloscopy and dactyloscopy: Do they dictate personality patterns?,” *J Forensic Dent Sci*, vol. 7, no. 2, pp. 114–120, 2015, doi: 10.4103/0975-1475.155079.

[17] A. D. B. Stamm, “Identificación humana a través de la queiloscopia:: diferencias de género mediante análisis de morfotipos de huellas labiales en Pirané, Argentina,” *Acta odontológica venezolana*, vol. 53, no. 3, pp. 23–24, 2015.

[18] P. Ramakrishnan, S. Bahirwani, and S. Valambath, “Assessment of cheiloscopy in sex determination using lysochrome - A preliminary study,” *J Forensic Dent Sci*, vol. 7, no. 3, pp. 195–200, 2015, doi: 10.4103/0975-1475.172434.

[19] R. Sheno, P. T. U. B. M. Gopalan, and J. Nadankutty, “Cheiloscopy – A Study of Lip Prints for Personal Identification,” *IOSR Journal of Dental and Medical Sciences*, vol. 15, pp. 101–103, Feb. 2016, doi: 10.9790/0853-1525101103.

BDDM – Bases de Datos y Minería de Datos

Análisis y Aplicaciones de Internet de las Cosas y Ciudades Inteligentes.

Escenario de Testeo de Seguridad

Gramajo Sergio, Scappini Reinaldo, Bolatti Diego, Todt Carolina

Universidad Tecnológica Nacional, Facultad Regional Resistencia
Departamento de Ingeniería en Sistemas de Información
Centro de Investigación Aplicada en Tecnologías de la Información
y la Comunicación (CInApTIC)
French 414 – Resistencia (3500) Chaco - Argentina
{sergiogramajo, rscappini, diegobolatti, carolinatodt}@gmail.com

RESUMEN

El rápido desarrollo de las nuevas tecnologías de Internet de las Cosas (IoT) como sus usos en diferentes campos de la industria, las ciudades, la salud, los hogares, entre otros, ha planteado nuevas formas de enfrentar amenazas de seguridad. IoT es una colección de dispositivos interconectados fortalecido con pequeños procesadores, placas o interfaces de red que se comunican con servicios web u otro tipo de interfaces a través de diferentes medios de telecomunicación.

Naturalmente si una nueva tecnología es ampliamente adoptada por el público y hay una notoria falta de estándares para el campo, las amenazas de ciberseguridad crecen.

En esta etapa del proyecto nos centramos en generar un escenario para realizar un relevamiento de los diferentes tipos de tráfico y crear un posterior datasets de análisis para poder identificar diferentes amenazas en sistemas IoT industriales o corporativos.

Palabras Clave: Internet de las cosas, Ciudades Inteligentes, Framework.

CONTEXTO

Este trabajo de investigación se desarrolla en el marco del proyecto “Análisis y Aplicaciones de Internet de las Cosas y Ciudades Inteligentes basadas en

Telecomunicaciones y Seguridad” del Centro de Investigación Aplicada en TICS (CInApTIC) de la Universidad Tecnológica Nacional, Facultad Regional Resistencia.

1. INTRODUCCIÓN

Como hemos mencionado, actualmente el Internet de las cosas (IoT) vive una gran expansión [1] y fue el resultado de la evolución sistemática de las telecomunicaciones y la electrónica de sensores [2] [3] acompañada con la demanda de la sociedad y dispositivos con mínima intervención humana [4]. En este contexto de intercomunicación, los sistemas pueden enviar información y tomar decisiones [5]. Para ello deben hacerlo en un ecosistema seguro y confiable y hoy existe un creciente número de amenazas de ciberseguridad para ellos [6].

Es así que, en este punto del proyecto de investigación se está desplegando un escenario de testeo (monitor) que pueda ser modelo para detectar tráfico anómalo y clasificar oportunamente amenazas o tráfico normal.

El diseño del monitor de tráfico se creó a partir de un bridge y un contenedor Docker con un software de captura tcpdump. Esta captura se realiza en modo “raw” obteniendo la totalidad de los paquetes de una determinada interfaz del puente al cual se conecta el contenedor monitor, sin privilegios elevados, solo un contenedor

convencional. El puente al reflejar todo el tráfico de la interfaz elegida permite además que múltiples herramientas obtengan los mismos datos, siendo esto de gran utilidad si por ejemplo se desea establecer funciones de selección y discriminación de tráfico, siendo éste un requisito fundamental para el desarrollo del proyecto.

Cabe destacar que el uso de docker facilita un entorno aislado y fácilmente replicable que asegura la portabilidad e implementación del monitor allí donde haga falta.

Las etapas que se están llevando a cabo para la creación del escenario son las siguientes:

1. Configuración de conexión VPN a la red de la UTN FRRe en terminales Windows10 y Ubuntu 20.04.
2. Inspección y verificación de la MV que da soporte al Gateway LORA y donde residen los servicios implementados en el ámbito del proyecto.
3. Instalación vía SSH del software requerido para el desarrollo de las actividades de investigación (soporte NetworManagerCLI, Docker, OpenVswitch, y otros).
4. Copiar los contenedores necesarios.
5. Montaje de escenario de pruebas
6. Ajuste del diseño desarrollado y probado
7. Rediseño conforme el contexto y soporte de la plataforma de la MV UTI CINAPTIC.

La propuesta final es que el detector de anomalías esté basado en redes definidas por software (SDN) [7] y se ubique en la capa de dispositivo de la arquitectura de IoT, tal como se muestra en la Figura 1.

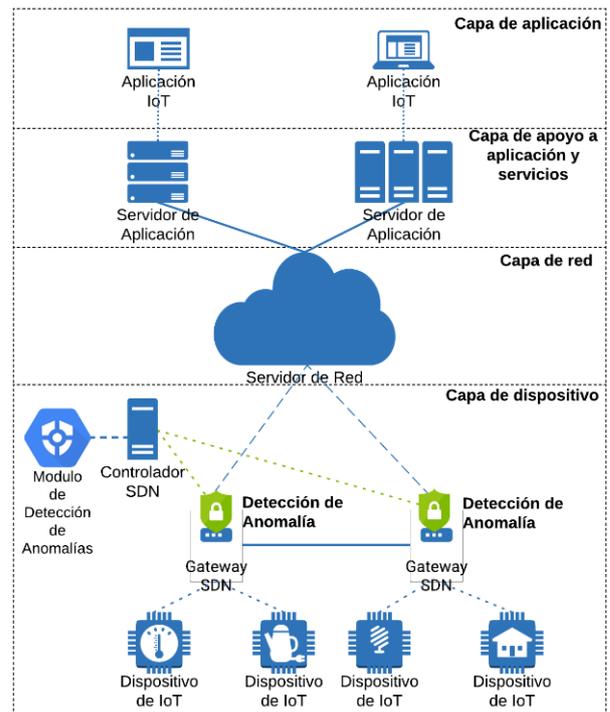


Figura 1: Arquitectura Propuesta.

En la Figura 1 se puede apreciar que el sistema está compuesto por:

- **Dispositivos finales de IoT:** de poca capacidad de procesamiento, como sensores de temperatura, luces inteligentes, entre otros.
- **Gateway SDN:** permiten que los dispositivos de IoT se conecten a la red. Para la conectividad entre los dispositivos finales y los Gateway se utilizan redes de baja potencia y área amplia (LPWAN) como LoRa, Sigfox, entre otros.
- **Controlador SDN:** este componente administra y configura los recursos de la red.
- **Módulo de detección de anomalías:** este módulo recopila los datos de las puertas de enlace para así buscar anomalías.

El rol de este módulo es agregar inteligencia al controlador SDN para reajustar la red y mantener las políticas de seguridad definidas por

los administradores al detectar anomalías.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las líneas de investigación que se abordan en el proyecto están vinculadas con:

- **Frameworks genérico de captación y presentación de datos:** Consiste en el desarrollo del framework que pueda ser adaptado a diferentes situaciones del medio local.
- **Arquitectura de redes de información para IoT:** Para el diseño de un detector de anomalías sobre IoT.
- **Inteligencia Artificial:** Se realizará un análisis de las técnicas de Machine Learning, con el objetivo de seleccionar la mejor opción para la implementación en el módulo de detección de anomalías.
- **Seguridad de IoT:** Se analizarán los ataques de seguridad y anomalías que se presentan en un entorno de IoT.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados obtenidos en el proyecto podrán ser utilizados en las siguientes áreas del conocimiento:

A) Arquitecturas de redes de Información para IoT. El relevamiento y análisis de las nuevas tendencias de redes de información de corto y amplio rango, conllevará a publicaciones científicas y transferencias al medio local o regional. Esto propiciará el contacto con investigadores de nivel internacional y nacional de otras instituciones para posibles intercambios de experiencias como el que se está llevando a cabo con el proyecto REMIND de la Unión Europea¹.

B) Programación y pruebas de diversos dispositivos usados para IoT y ciudades inteligentes como sensores y equipos de

telecomunicación entre ellos sin intervención humana y que ayude a la toma de decisiones y mejore la gestión que lo utilice.

Una vez finalizado el proyecto generará nuevo conocimiento y aplicaciones que pueden ser transferidos tanto a entornos de investigación como diferentes entornos organizacionales o empresas del medio. En este sentido se pretende impulsar el intercambio de conocimiento con investigadores de otras instituciones, asistencia a eventos científicos, elaboración de publicaciones científicas, estancias de investigación en el exterior, convenios de transferencia, etc.

4. FORMACIÓN DE RECURSOS HUMANOS

Con el proyecto se pretende contribuir a la formación de recursos humanos desde diversas áreas:

- **Formación de becarios:** El proyecto cuenta con la participación de alumnos becarios del último año de la carrera de Ingeniería en Sistemas de Información que están realizando su práctica supervisada.
- **Alumnos de la carrera de Ingeniería en Sistemas de Información:** Se prevé realizar actividades de actualización y talleres con alumnos de las cátedras del área de redes de información, comunicaciones y seguridad informática. Además, por las propias características de los temas que involucra el proyecto se pueden realizar actividades en cátedras como inteligencia artificial.
- **Formación de jóvenes profesionales:** Se prevé la incorporación de jóvenes profesionales de Ingeniería en Sistemas de Información con la intención de seguir con una carrera en investigación universitaria. Los

¹ (REMIND) de Horizonte2020 <https://remind-research.com/>

cuales pueden incorporarse en carácter ad-honorem al proyecto o a través de becas de iniciación en la investigación.

- **Formación de postgrado:** A partir de las líneas de investigación desarrolladas en el proyecto se prevé que el Ing. Diego Bolatti finalice su doctorado mediante una tesis vinculada a este proyecto.
- **Equipo de trabajo:**
 - **Director:**
 - Gramajo, Sergio.
 - **Investigadores de apoyo:**
 - Bolatti, Diego
 - Scappini, Reinaldo
 - **Becario alumno:**
 - Todt, Carolina
 - Federico Aguirre

5. BIBLIOGRAFÍA

[1] The number of smart homes in Europe and North America reached 45 million in 2017, 2018, IoT Business News, <https://iotbusinessnews.com/2018/09/24/20413-the-number-of-smart-homes-in-europe-and-north-america-reached-45-million-in-2017/>. Online. (Accessed 24 Feb 2022).

[2] Silva, B. N., Khan, M., & Han, K. Internet of things: A comprehensive review of enabling technologies, architecture, and challenges. IETE Technical Review, 1–16. 2017.

[3] Silva, B. N., Khan, M., & Han, K. Big data analytics embedded Smart City architecture for performance enhancement through real-time data processing and decision-making. Wireless Communications and Mobile Computing. 2017.

[4] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. Internet of Things (IoT): A vision, architectural elements, and future directions. Future Generation Computer Systems, 29, 1645–1660. 2013.

[5] Bhagya Nathali Silva, Murad Khan, Kijun Han. Towards sustainable Smart Cities: A review of trends, architectures, components, and open challenges in smart cities. Sustainable Cities and Society. 38. 2018

[6] N. Moustafa, B. Turnbull, K.-K.R. Choo, Towards automation of vulnerability and exploitation identification in iot networks, in: 2018 IEEE International Conference on Industrial Internet, ICII, IEEE, 2018, pp. 139–145.

[7] T. D. Nadeau y K. Gray, SDN: Software Defined Networks, 1ra ed. .O'Reilly Media, Inc.", 2013.

Modelado de Variedad de Activos de Dominio en Sistemas Big Data

Agustina Buccella, Alejandra Cechich, Juan Luzuriaga
Líam Osycka, Carolina Villegas, Marcos Cruz, Franco Corgatelli,
Rodolfo Martínez, Rafaela Mazalu, Marcelo Moyano
GIISCO Research Group
Departamento de Ingeniería de Sistemas
Universidad Nacional del Comahue
Neuquen, Argentina
agustina.buccella@fi.uncoma.edu.ar

1. Resumen

Un cambio importante con respecto a depósitos de datos tradicionales, es que en los Sistemas Big Data (SBDs) la naturaleza no estructurada de algunos datos puede provenir de diversas fuentes, entre ellas sensores, redes sociales, entorno y la misma empresa. La diversidad de esos datos puede analizarse abordando distintas características. Precisamente, la propiedad de los SBDs con respecto a diversidad de los datos se denomina *Variedad*.

La variedad en SBDs ha sido relacionada con diversas propiedades como interoperabilidad, seguridad, reusabilidad, etc. En este contexto, y respondiendo a la pregunta de investigación: *¿Cómo puede modelarse la variedad de la información de dominio de manera de incorporar reusabilidad en el desarrollo de SBDs?*, nuestro proyecto propone modelar variedad a modo de líneas de productos. A diferencia de otras propuestas, la nuestra toma como partida una estructura de actividades asociadas al desarrollo de SBDs, instanciada en artefactos software producidos durante esas actividades e incorpora el modelado de variedades de manera similar a líneas de productos.

Palabras Clave: Reusabilidad - Líneas de Producto de Software - Big Data

2. Contexto

La línea presentada se inserta en el contexto del *Proyecto UNComa: Modelado de Variedad en Sistemas Big Data*. Directora: Dra. Agustina Buccella, Co-directora; Dra. Alejandra Cechich, que se encuentra en etapa de evaluación (2022-2025).

3. Introducción

En [1], la *Variedad* se clasifica en una taxonomía que divide el análisis en cuatro casos de diversidad: estructural, de las fuentes, de contenido y de procesamiento. Por ejemplo, la diversidad estructural denota la variedad en formatos y tipos de datos, clasificándolos como estructurados, semi-estructurados y no estructurados; la diversidad de las fuentes se clasifica en tres grupos - datos generados por humanos, generados por máquinas o mediados por procesos; la diversidad de contenido aborda diferentes tipos de soporte (único medio, multimedia o gráfico); y la diversidad de procesamiento enfoca en las distintas necesidades de procesamiento algorítmico (batch, interactivo, streaming o gráfico).

La variedad en los datos también ha sido considerada desde el punto de vista de incorporación de semántica al proceso de modelado de arquitecturas en SBDs; por ejemplo, en

[2] se utiliza modelado de contextos para mejorar las aplicaciones Big Data en el ámbito de ciudades inteligentes y a efectos de analizar su sustentabilidad. En [11], en cambio, se sugiere el uso de técnicas de Deep Learning para extraer patrones complejos en los datos, así como para permitir indexación semántica.

Por otra parte, la variedad en SBDs ha sido relacionada con diversas propiedades como interoperabilidad, seguridad, reusabilidad, etc. Por ejemplo, en [9], se presentan seis arquitecturas de referencia propuestas actualmente en la literatura para SBDs y se analiza el cumplimiento de las cinco Vs (Volumen, Velocidad, Variedad, Variabilidad y Veracidad) como requerimientos tradicionales. En el análisis, la variedad se asocia a los siguientes requerimientos: a (R1) la heterogeneidad de los datos en formatos no estructurados (ej. texto y video), semi-estructurados (ej. basados en XML o JSON) y estructurados (ej. tablas relacionales); a (R2) el análisis eficiente de los datos lo cual requiere entender exactamente qué está almacenado y a (R3) resolver conflictos de interoperabilidad, lo que implica el manejo de metadatos apropiados. R2 puede abordarse por medio de métodos para integración de datos (ej. transformación basada en reglas), que permiten resolver conflictos en los esquemas de datos. Sin embargo, R3 requiere de una efectiva integración de los datos para resolver conflictos de interoperabilidad: (1) conflictos de dominio que se relacionan con diferentes interpretaciones del mismo dominio, incluyendo homónimos, sinónimos, acrónimos y restricciones de integridad; (2) conflictos de granularidad que se relacionan con diferentes unidades de medida y agregación de los datos; y (3) conflictos de completitud que se relacionan con diferentes piezas de datos que pertenecen a la misma entidad. De las comparaciones realizadas, se desprende que casi todas las arquitecturas de referencia propuestas satisfacen el requerimiento R1, pero sólo algunas abordan parcialmente R2 y ninguna al momento satisface R3, dejando un amplio margen para aportes en el área.

En SBDs, la reusabilidad ha sido abordada también desde diversos ángulos. Por ejemplo, en [13] se discuten conceptos de reusabi-

lidad en el contexto de analítica de datos distinguiendo entre uso y reuso del dato. Más específicamente, pero en el mismo sentido, en [8] se profundizan aspectos de privacidad en el contexto de reusabilidad de datos. Allí se propone una taxonomía en reuso de datos que pueda ser útil para determinar en qué medida ese reuso debe ser permitido y bajo qué condiciones para preservar privacidad. Otras propuestas, abordan aspectos de reuso en términos de aumentar la colaboración en el desarrollo de SBDs mediante el uso de nuevas tecnologías (ej. computación en la nube). Por ejemplo, en [15] se propone un enfoque de gestión de SBDs mediante capacidades de almacenamiento y procesamiento en una nube pública y se ejemplifica su uso. Adicionalmente, las distintas plataformas de soporte para desarrollo de SBDs también se abordan desde un punto de vista de reuso; por ej. en [3] se analiza la mejora en la eficiencia de herramientas como Apache Hadoop¹ y Spark² debido al reuso de artefactos entre diversos proyectos. Para ello, se analizan aspectos comunes y se provee de un flujo de trabajo implementado de manera escalable y extensible.

En un sentido similar, incorporando la detección de aspectos comunes y variables a modo de familia de sistemas, en [10] la arquitectura de referencia se ve acotada por medio de casos de uso (ej. visualización y análisis de información geoespacial estratégica, análisis inteligente de señales, etc.). De esos casos, se identifican requerimientos relevantes al SBD, incluyendo categorías, como tipos de datos (ej. texto no estructurado, geoespacial, audio), transformaciones en los datos (ej. clustering, correlación), visualizaciones (ej. imágenes, redes), etc. Luego, la arquitectura se organiza como una colección de módulos que descomponen la solución en elementos realizando funciones o capacidades para un conjunto de aspectos (requerimientos externos, módulos reusables, roles y participantes, paquetes de datos comerciales y open source). Finalmente, se descompone el SBD en 13 módulos agrupados en las categorías: (1) Proveedor de aplicaciones de Big Data (incluyendo la ló-

¹<https://hadoop.apache.org/>

²<https://spark.apache.org/>

gica de negocios del sistema), (2) Proveedor del framework de Big Data (incluyendo plataformas, almacenamiento, etc.), y (3) Módulos transversales (abordando distintos aspectos relevantes al desarrollo de SBDs).

Considerando estas propuestas, en [6] hemos definido una primera aproximación de los elementos que componen nuestra arquitectura de referencia para SBDs basada en reuso. En ella, los aspectos de negocios (dominio), aplicación (software y análisis) y tecnológicos se abordan en niveles separados; siendo transversales aspectos como el uso/reuso de estándares, taxonomías y conocimiento.

Uno de los componentes principales de esta arquitectura agrupa los denominados *activos de dominio*, constituidos por artefactos de software que son creados para el dominio en el que se está trabajando. Así, además de incluir a los participantes del desarrollo del SBD, involucra los requerimientos del proyecto y del dominio, restricciones, modelos y casos de uso. Es importante resaltar que estos activos deben generarse a partir de *taxonomías de dominio y estándares* y de *activos basados en conocimiento*. De esta forma, se deben crear artefactos enfocados en que puedan ser reusados en el mismo dominio e incluso en otros dominios relacionados (artefactos para reuso), y/o que puedan desarrollarse en base a otros artefactos ya creados (artefactos con reuso).

Para identificar variedad en activos de dominio, proponemos dos enfoques: (1) identificación desde los datos (bottom-up) y/o identificación desde los requerimientos (top-down). Ambos casos no son excluyentes, ya que se puede iniciar una búsqueda a modo exploratorio desde los datos, a la vez que se establecen algunos requerimientos de búsqueda.

En la Figura 1 mostramos la visión global del enfoque bottom-up de nuestra propuesta [12], es decir, la identificación de variedad a partir de los datos (Paso (1) “¿Qué dicen los datos?”). Se procede luego a realizar el estudio exploratorio (Paso (2)), que brindará información de correlaciones posibles en los datos y variaciones detectadas en su análisis.

En principio, al centrarnos en SBDs, el primer elemento a considerar es el proceso de desarrollo (Paso (3)), donde las etapas bási-

cas pueden resumirse en: (1) Adquisición de Datos, que consiste en extraer los datos desde las fuentes, agregando un proceso de carga y filtrado para que los datos sean adecuados a su posterior procesamiento; (2) Preparación de Datos, que consiste en estructurar el formato de los datos, realizar la limpieza de los mismos y eventualmente, también su integración; y (3) Análisis de Datos, que contiene las funcionalidades que permiten derivar conocimiento a partir de los datos, enfocando en análisis descriptivo, predictivo y/o prescriptivo.

Luego, los resultados validados con los expertos (Paso (4)) podrán utilizarse en decisiones referidas al problema de dominio (Paso (5)), e incluso retroalimentar un nuevo ciclo exploratorio.

Entonces, para identificar variedad en los activos de dominio, el enfoque bottom-up de la propuesta parte de la definición de un problema dependiente del dominio e intenta detectar características variantes dentro de cada una de las etapas del proceso de análisis de datos (Paso (3)). En el ejemplo de la Figura 1, sólo se enfoca en identificar la diversidad de contexto (o dominio), manteniendo constantes las fuentes, contenido y procesamiento.

Al igual que en el desarrollo en líneas de producto software (LPSs), nuestra propuesta modela la variedad en activos de dominio a través de dos fases: la Ingeniería de Dominio y la Ingeniería de Aplicación. La primera es responsable de identificar y definir las funcionalidades y aspectos comunes y variables que forman parte de la plataforma que comparten todos los productos de la línea de productos software (todos los SBDs); mientras que la segunda está compuesta por las actividades que permiten realizar la derivación de productos particulares, es decir, que permiten instanciar en cada caso [14].

En trabajos previos, hemos presentado una propuesta de diseño de LPSs dirigida por funcionalidades, donde cada funcionalidad se documenta a través de una hoja de datos funcional (datasheet), representando el conjunto de servicios comunes y variantes [4, 5]. Para el caso de reusabilidad en SBDs, la Figura 1 muestra la hoja de datos funcional definida para reusar modelos de análisis de datos (activos

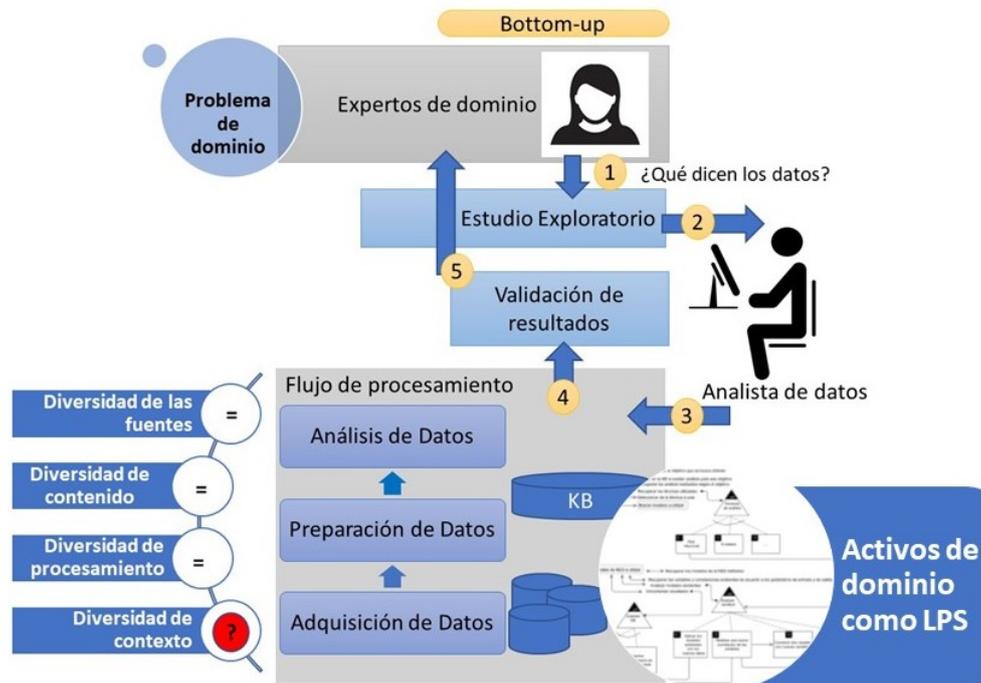


Figura 1. Vision global del enfoque bottom-up

de dominio a nivel de plataforma de LPS) y detectar variedades de contexto. Detalles de este modelo pueden verse en [12].

4. Líneas de Investigación y Desarrollo

En proyectos previos, hemos realizado amplios avances en lo que respecta al área de LPSs definiendo y refinando una metodología de desarrollo a nivel de subdominios. Dentro de la metodología, hemos presentado sus bases y diseñado artefactos que se utilizan en el análisis de dominios y en el análisis organizacional de una LPS [4] y tienen la particularidad de favorecer el reuso basado en una taxonomía de servicios. Es precisamente esta ventaja la que nos permitió luego realizar extensiones hacia otros subdominios. De esta forma hemos podido así avanzar en el desarrollo de múltiples LPSs basadas en la jerarquía de dominios definida. Sin embargo, para dicha extensión hemos tenido que formalizar varios aspectos respecto a los artefactos de software creados. En el caso de la taxonomía, se utilizaron los estándares reconocidos del subdominio, y se creó un proceso de desarrollo que se aplicó en la creación de las dos primeras LPSs [5, 7].

5. Resultados Obtenidos/Esperados

El objetivo principal de la línea de investigación es *Desarrollar técnicas y herramientas que mejoren los procesos y técnicas aplicadas a la explotación de grandes volúmenes de datos, favoreciendo el desarrollo de ambientes inteligentes que permitan reusabilidad.*

Al momento, hemos planteado una arquitectura de referencia [6] y un modelo de procesos [12] para el modelado y gestión de la variedad en SBDs. Los resultados publicados son preliminares a modo de prueba de conceptos. Actualmente, estamos trabajando en colaboración con el Instituto de Tecnología Agropecuaria (INTA)-Alto Valle para la aplicación del proceso de modelado en el análisis de la napa freática, en función de la variedad de fuentes acuíferas de diversas zonas geográficas (variedad contextual).

6. Formación de Recursos Humanos

El proyecto reúne aproximadamente a 13 investigadores, entre los que se cuentan docentes y alumnos de UNComa, y colaboradoras expertas del dominio de aplicación, espe-

cíficamente pertenecientes al INTA. A su vez, el proyecto cuenta actualmente con dos doctores y un magister. Varios de los docentes-investigadores de GIISCo-UNComa han terminado o se encuentran próximos a terminar carreras de postgrado. Además, varios de los integrantes se encuentran finalizando sus tesis de grado. Por último, este año seguiremos con la supervisión del trabajo de 2 becarios EVC-CIN.

Referencias

- [1] Jemal Abawajy. Comprehensive analysis of big data variety landscape. *International Journal of Parallel, Emergent and Distributed Systems*, 30(1):5–14, 2015.
- [2] S. Bibri and J. Krogstie. The core enabling technologies of big data analytics and context-aware computing for smart sustainable cities: a review and synthesis. *Journal of Big Data*, 4(38), 2017.
- [3] Reuben Borrison, Benjamin Klöpper, Moncef Chioua, Marcel Dix, and Barbara Sprick. Reusable big data system for industrial data mining - a case study on anomaly detection in chemical plants. In *Intelligent Data Engineering and Automated Learning – IDEAL 2018*, pages 611–622. Springer International Publishing, 2018.
- [4] A. Buccella, A. Cechich, M. Arias, M. Pol’la, S. Doldan, and E. Morsan. Towards systematic software reuse of gis: Insights from a case study. *Computers & Geosciences*, 54(0):9 – 20, 2013.
- [5] A. Buccella, A. Cechich, M. Pol’la, M. Arias, S. Doldan, and E. Morsan. Marine ecology service reuse through taxonomy-oriented SPL development. *Computers & Geosciences*, 73(0):108 – 121, 2014.
- [6] A. Buccella, J. Luzuriaga, A. Cechich, L. Osycka, F. Paterno, M. Pol’la, M. Cruz, R. Martinez, R. Mazalu, and M. Moyano. Reusabilidad en el contexto de desarrollo de sistemas para big data. In *Actas del XXIII Workshop de Investigadores en Ciencias de la Computación, Chilecito, La Rioja*, pages 525–529, 2021.
- [7] Agustina Buccella, Alejandra Cechich, Juan Porfiri, and Domenica Diniz Dos Santos. Taxonomy-oriented domain analysis of gis: A case study for paleontological software systems. *ISPRS International Journal of Geo-Information*, 8(6), 2019.
- [8] Bart Custers and Helena Uršič. Big data and data reuse: a taxonomy of data reuse for balancing big data benefits and personal data protection. *International Data Privacy Law*, 6(1):4–15, 2016.
- [9] Ali Davoudian and Mengchi Liu. Big data systems: A software engineering perspective. *ACM Computing Surveys*, 53(5), 2020.
- [10] John Klein, Ross Buglak, David Blockhow, Troy Wuttke, and Brenton Cooper. A reference architecture for big data systems in the national security domain. In *Proceedings of the 2nd International Workshop on BIG Data Software Engineering*. ACM/IEEE, 2016.
- [11] M. Najafabadi, F. Villanustre, T. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic. Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1), 2017.
- [12] L. Osycka, A. Buccella, and N. A. Cechich. Identificación de variedad contextual en modelado de sistemas big data. In *Memorias del XXVII Congreso Argentino de Ciencias de la Computación (CACIC)*, pages 367–376. Red de Universidades con Carreras en Informática, 2021.
- [13] I.V. Paschetto, B.M. Randles, and C.L. Borgman. On the reuse of scientific data. *Data Science Journal*, 16(8), 201720.
- [14] Klaus Pohl, Günter Böckle, and Frank J. van der Linden. *Software Product Line Engineering: Foundations, Principles and Techniques*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.
- [15] Zhiwu Xie, Yinlin Chen, Julie Speer, Tyler Walters, Pablo A. Tarazaga, and Mary Kasarda. Towards use and reuse driven big data management. In *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries*, page 65–74. Association for Computing Machinery, 2015.

ABORDAJE METODOLÓGICO DE MINERÍA DE DATOS APLICADA A INVESTIGACIONES AGROPECUARIAS SOBRE ARTRÓPODOS

Mariano R. Droz ^(1,2), Carlos E. Alvez ⁽¹⁾, Pedro D. Benitez ⁽²⁾, Juan A. Ramos ⁽²⁾,
Beatriz M. Diaz ⁽³⁾, Luz M. Zapata ⁽²⁾

⁽¹⁾ Facultad de Ciencias de la Administración – Universidad Nacional de Entre Ríos
Av. Monseñor Tavella 1424 – (CP 3.200) Concordia, Entre Ríos, República Argentina

⁽²⁾ Facultad de Ciencias de la Alimentación – Universidad Nacional de Entre Ríos
Av. Monseñor Tavella 1450 – (CP 3.200) Concordia, Entre Ríos, República Argentina

⁽³⁾ Estación Experimental Agropecuaria Concordia – Instituto Nacional de Tecnología Agropecuaria
Ruta Provincial 22 y vías del Ferrocarril – (CP 3.200) Estación Yuquerí, Concordia, Entre Ríos,
República Argentina

{*mariano.droz, carlos.alvez, pedrodaniel.benitez, juan.ramos, luzmarina.zapata*}@uner.edu.ar
diaz.beatriz@inta.gob.ar

RESUMEN

La Minería de Datos (*Data Mining*) trata de resolver problemas o comprender fenómenos o situaciones mediante el análisis de datos digitales. Minería de Datos y extracción o “descubrimiento” de conocimiento en bases de datos (*Knowledge Discovery in Databases, KDD*) se han empleado indistintamente, pero existen diferencias entre ambas, siendo la primera solamente el núcleo de un proceso KDD. No obstante, actualmente Ciencia de Datos (*Data Science*) es una expresión mucho más utilizada en el contexto de descubrimiento de conocimiento a partir de datos, por eso en este trabajo se aborda un proyecto de Minería de Datos con un enfoque de ciencia de datos aplicada. Llevar a cabo un proyecto de este tipo requiere de un abordaje metodológico, se ha optado por CRISP-DM (*Cross Industry Standard Process for Data Mining*) ya que es considerado como el estándar de facto para proyectos de analítica, Minería de Datos y Ciencia de Datos. El objetivo consiste en ejecutar un proyecto para descubrir, determinar o relacionar la incidencia de los factores abióticos (temperatura y humedad relativa ambiente, nivel de luz, y temperatura y humedad del suelo) en el comportamiento de la artropofauna edáfica existente en sistemas productivos hortícolas de la región de Salto Grande.

Palabras clave: *Bases de Datos, Minería de Datos, Ciencia de Datos, CRISP-DM, Artrópodos.*

CONTEXTO

Este trabajo es desarrollado en el marco de las actividades de la Tesis denominada “Minería de Datos aplicada a estudios de biodiversidad de artrópodos de suelo” de la Maestría en Sistemas de Información (Facultad de Ciencias de la Administración. Universidad Nacional de Entre Ríos. Resolución “C.D.” N° 144/20). A su vez, está vinculado al Proyecto PID Novel N° 8112 “Diseño y desarrollo de una trampa de caída por tiempo con sensores y datalogger para estudios de biodiversidad de artrópodos de suelo” de la Universidad Nacional de Entre Ríos (UNER).

Por otra parte, se enmarca en la línea de trabajo “Desarrollo de tecnologías de bajo impacto ambiental aplicadas a la horticultura”, que lleva a cabo el Grupo Hortícola de la Estación Experimental Agropecuaria Concordia (EEA Concordia) del Instituto Nacional de Tecnología Agropecuaria (INTA).

1. INTRODUCCIÓN

De acuerdo con [1] la Minería de Datos (*Data Mining, DM*) trata de resolver problemas

mediante el análisis de datos existentes en bases de datos. A través del tiempo, varias expresiones se han utilizado para hacer referencia a la Minería de Datos, entre ellas: descubrimiento de conocimiento en bases de datos (*Knowledge Discovery -mining- in Databases, KDD*), extracción de conocimiento (*knowledge extraction*), análisis de patrones (*data/pattern analysis*), arqueología de datos (*data archeology*) y recolección de información (*information harvesting*) [2, 3].

Si bien Minería de Datos y extracción o “descubrimiento” de conocimiento en bases de datos (KDD) se han empleado indistintamente, existen diferencias entre ambas. El término KDD se utiliza para referirse a un proceso que consta de una serie de fases, mientras que la Minería de Datos es sólo una de estas fases [4].

Así, la Minería de Datos es el núcleo de un proceso KDD, ya que implica aplicar algoritmos para explorar los datos, desarrollar modelos y descubrir patrones previamente desconocidos. Estos modelos resultantes son utilizados para analizar, comprender o predecir fenómenos a partir de los datos [5].

Sin embargo, actualmente Ciencia de Datos (*Data Science*) es una expresión mucho más utilizada que Minería de Datos en el contexto de descubrimiento de conocimiento a partir de datos [6].

Ahora bien, ¿qué es la Ciencia de Datos?, según [6] existen dos perspectivas: a) ciencia de los datos; y b) aplicar métodos científicos a los datos. Para estos autores estos dos enfoques podrían llamarse: a) ciencia de datos teórica; y b) ciencia de datos aplicada.

Desde el punto de vista de la primera perspectiva, se hace un abordaje académico, se estudian los datos en todas sus manifestaciones, junto con los métodos y algoritmos para manipular, analizar, visualizar y enriquecer los datos. Esta perspectiva es metodológicamente cercana a la informática y a la estadística, combinando trabajo teórico, algorítmico y empírico [6].

En tanto, desde la segunda perspectiva, la ciencia de datos abarca tanto el enfoque académico como el industrial, extrayendo valor

de los datos utilizando métodos científicos, como las pruebas de hipótesis estadísticas o el aprendizaje automático. Aquí el énfasis está en resolver problemas específicos de un dominio en función de los datos. Éstos se utilizan para construir modelos, diseñar artefactos y, en general, aumentar la comprensión de un tema [6].

A partir de lo anterior, en este trabajo se hace referencia a un proyecto de Minería de Datos como un proceso de extracción de conocimiento a partir de datos, lo que también se considera ciencia de datos aplicada.

Un proyecto de Minería de Datos de este tipo es recomendable que sea abordado siguiendo una metodología. Así, CRISP-DM [7] (*Cross Industry Standard Process for Data Mining*), presentada en el año 1999 por las empresas NCR, SPSS y Daimler Chrysler, es actualmente considerado el estándar de facto para proyectos de analítica, Minería de Datos y Ciencia de Datos [6].

CRISP-DM elabora y amplía las fases de la propuesta original de KDD en seis fases: Comprensión o entendimiento del negocio (*Business understanding*), Comprensión de los datos (*Data understanding*), Preparación de los datos (*Data preparation*), Modelización (*Modelling*), Evaluación (*Evaluation*) y Despliegue (*Deployment*). Cada fase se descompone en un conjunto de tareas genéricas (o generales) de segundo nivel. A partir del tercer nivel de abstracción, se realiza un “mapeo” de las tareas genéricas definidas en el modelo a situaciones específicas. En el cuarto nivel, se encuentran las instancias de proceso, donde se describen las acciones, decisiones y resultados de un proyecto particular de Minería de Datos [6, 8].

En cuanto al campo de uso de la Minería de Datos, es imposible efectuar una enumeración pormenorizada de todas las aplicaciones donde juega un papel crítico, ya que donde hay datos, hay aplicaciones [2]. Más allá de esto, específicamente este trabajo apunta a llevar a cabo un proyecto de Minería de Datos para extraer conocimiento con el fin de descubrir o explicar patrones de comportamiento de

artropodos de suelo a partir de ciertos factores abióticos.

Cabe señalar, que en los ecosistemas terrestres la fauna edáfica juega un papel clave en la provisión de funciones (ciclado de nutrientes, descomposición de la materia orgánica, mantención de hábitats para organismos benéficos, etc.) y servicios ecosistémicos muy valorados en la actualidad como componentes de la biodiversidad funcional [9, 10].

Una de las formas de realizar los estudios de la biodiversidad de los artrópodos epigeos del suelo es mediante la utilización de trampas de caída, llamadas habitualmente “pitfall” [11, 12]. El avance tecnológico ha permitido la evolución de las trampas pitfall, surgiendo así las trampas de caída por tiempo (*time-sorting pitfall trap*) y las trampas de caída por tiempo con sensores [11–14].

El presente trabajo da continuidad a la cooperación científico-tecnológica entre dos grupos de investigación de la Universidad Nacional de Entre Ríos y la Estación Experimental Agropecuaria Concordia del Instituto Nacional de Tecnología Agropecuaria. Los datos recolectados en ocho ensayos de campo realizados durante el año 2021, utilizando una trampa automatizada de caída por tiempo con sensores y datalogger, diseñada y construida para estudios de biodiversidad de artrópodos de suelo, serán integrados a una base de datos con el propósito de llevar a cabo un proyecto de Minería de Datos guiado metodológicamente por CRISP-DM.

De esta manera, el aporte de esta propuesta consiste en aplicar Minería de Datos en una base de datos específicamente diseñada para descubrir, determinar o relacionar la incidencia de los factores abióticos (temperatura y humedad relativa ambiente, nivel de luz, y temperatura y humedad del suelo) en el comportamiento de la artropofauna edáfica existente en sistemas productivos hortícolas de la región de Salto Grande.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El presente trabajo se enmarca en el análisis y aplicación de Minería de Datos con fines científicos para estudios de biodiversidad de artrópodos de suelo.

Contempla el diseño, implementación y carga de datos en una base de datos específica para este propósito. A su vez, comprende aplicar la metodología CRISP-DM para guiar el proceso de extracción de conocimiento sobre patrones de comportamiento de artrópodos de suelo, a partir de datos genuinos obtenidos en ensayos de campo realizados con una trampa “pitfall” convencional y otra automatizada por tiempo con sensores de temperatura y humedad relativa ambiente, nivel de luz y temperatura y humedad del suelo.

Asimismo, esta propuesta se encuadra en líneas de investigaciones prioritarias de dos facultades de la Universidad Nacional de Entre Ríos. Por un lado, en la línea de investigación “*Bases de Datos*” de la Facultad de Ciencias de la Administración, según las Resoluciones “C.D.” N° 160/11 y 203/11. Por otro lado, también se inscribe en la línea de investigación “*Bioinformática*” correspondiente a la Facultad de Ciencias de la Alimentación, de acuerdo a la Resolución “C.D.” N° 313/19.

3. RESULTADOS ALCANZADOS Y ESPERADOS

El objetivo general del Proyecto PID Novel N° 8112 “Diseño y desarrollo de una trampa de caída por tiempo con sensores y datalogger para estudios de biodiversidad de artrópodos de suelo” fue diseñar y desarrollar una trampa de caída por tiempo automatizada para capturar artrópodos de suelo y medir factores abióticos con el fin de evaluar tecnologías de bajo impacto ambiental aplicadas a la horticultura.

Dicho objetivo general y los objetivos específicos vinculados al diseño, construcción, comprobación de su funcionamiento a través de ensayos de campo y evaluación de la capacidad de captura del dispositivo desarrollado, ya fueron alcanzados.

A partir de lo anterior, y del trabajo de acondicionamiento y clasificación de los individuos que componen las muestras, realizado por profesionales del Grupo Hortícola de la EEA Concordia del INTA, se disponen de datos propios que serán la materia prima principal sobre la cual se podrá llevar a cabo este proyecto de Minería de Datos.

El proyecto precitado será guiado por la metodología CRISP-DM y tiene como objetivo general descubrir conocimiento válido, útil y novedoso en una base de datos diseñada para estudios de biodiversidad de artrópodos de suelo.

En función a lo anterior, entre los objetivos específicos se encuentra diseñar e implementar en un motor de libre distribución una base de datos para unificar los datos obtenidos de orígenes y formatos diferentes (archivos del datalogger, planillas de cálculo, etc.). A su vez, el proyecto apunta a la construcción de modelos descriptivos a partir de los datos, con el propósito de extraer conocimiento sobre el comportamiento de la artropofauna edáfica estudiada.

Para dar cumplimiento a ello, en primer lugar se efectuará una búsqueda, revisión, clasificación y análisis de libros, trabajos y documentación científica y tecnológica relacionados con la temática. El foco de atención principal estará puesto en el proceso de extracción de conocimiento en bases de datos y en las tareas, los modelos y las técnicas de Minería de Datos.

Asimismo, se tiene previsto llevar a cabo y documentar las salidas de las tareas que prevé CRISP-DM en cada una de las seis fases, a saber [6, 7, 15]:

- Comprensión del problema o negocio: Determinar los objetivos del negocio, evaluar la situación, determinar los objetivos de la Minería de Datos y crear un plan para el proyecto.
- Entendimiento de los datos: Recolectar los datos iniciales, describir los datos, explorar los datos y verificar la calidad de los datos.

- Preparación de los datos: Seleccionar, limpiar, construir, integrar y dar forma a los datos.
- Modelado: Seleccionar la técnica de modelado, diseñar las pruebas del modelo, construir y evaluar el modelo.
- Evaluación: Evaluar los resultados, revisar el proceso y determinar las próximas etapas.
- Implementación: Planificar la implementación, planificar el monitoreo y el mantenimiento, crear un reporte final y revisar el proyecto.

Por consiguiente, la contribución del trabajo radica en que es una continuidad de una labor ya iniciada, la cual contempló desde la construcción del dispositivo, la realización de los ensayos de campo para obtener datos genuinos y, finalmente, la realización de un proyecto de Minería de Datos, sobre una base de datos surgida del trabajo en colaboración de investigadores de dos instituciones públicas de la República Argentina.

4. FORMACIÓN DE RECURSOS HUMANOS

Este trabajo forma parte de las actividades planificadas para una Tesis de la Maestría en Sistemas de Información de la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos.

El equipo de trabajo está compuesto por el tesista, Director y Co-Directora de la Tesis precitada y por el Director e investigadores del Proyecto PID Novel N° 8112 de la Facultad de Ciencias de la Alimentación de la Universidad Nacional de Entre Ríos. Asimismo, forma parte la investigadora responsable del Grupo Hortícola de la EEA Concordia del INTA.

A su vez, este trabajo es un ámbito idóneo para que estudiantes de carreras de grado afines de la Universidad Nacional de Entre Ríos, puedan llevar a cabo sus proyectos de Trabajo Final o participen como Becarios de Iniciación en la Investigación.

5. BIBLIOGRAFÍA.

1. Witten H. Ian, Frank Eibe, H.A.M.: Data Mining: Practical Machine Learning Tolls and Techniques. Elseiver Inc. (2011)
2. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques. (2011)
3. Vasilakos, A. V., Wan, J., Chen, F., Rong, X., Zhang, D., Deng, P.: Data Mining for the Internet of Things: Literature Review and Challenges. *Int. J. Distrib. Sens. Networks.* 11, 431047 (2015). <https://doi.org/10.1155/2015/431047>
4. Orallo, J.H., Quintana, M.J.R., Ramírez, C.F.: Introducción a la Minería de Datos. Pearson Educación (2004)
5. Maimon, O., Rokach, L.: The Data Mining and Knowledge Discovery Handbook.
6. Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M.J., Flach, P.: CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Trans. Knowl. Data Eng.* 33, 3048-3061 (2021). <https://doi.org/10.1109/TKDE.2019.2962680>
7. Pete, C., Julian, C., Randy, K., Thomas, K., Thomas, R., Colin, S., Wirth, R.: Crisp-Dm 1.0. *Cris. Consort.* 76 (2000)
8. Moine J.: Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo. XVII Congr. Argentino Ciencias La Comput. XVII CACIC, 15, 18-22 (2013)
9. ELN-FAB: Functional agrobiodiversity: Nature serving Europe's farmers. ECNC-European Centre for Nature Conservation, Tilburg, the Netherlands (2012)
10. Ruiz-Lupi3n, D., Pascual, J., Melguizo-Ruiz, N., Verdeny-Vilalta, O., Moya-Lara3o, J., Activity, S.A.: New litter trap devices outperform pitfall traps for studying arthropod activity. *Insects.* 10, 1-15 (2019). <https://doi.org/10.3390/insects10050147>
11. Droz, M.R., Diaz, B.M., Ramos, J.A., Benitez, P.D., Zapata, L.M.: Trampa pitfall con embudo y data logger para estudios de biodiversidad de artr3podos de suelo. *An. CAI 2020. Congr. ARGENTINO AGROINFORMATICA (JAIIO).* ISSN 2525-0949. 354-367 (2020)
12. Droz, M.R., Ramos, J.A., Benitez, P.D., Zapata, L.M., Diaz, B.M.: Dise3o y desarrollo de un sistema embebido para una trampa pitfall con data logger. XXVI Congr. Argentino Ciencias la Comput. (CACIC 2020). 571-580 (2021)
13. Buchholz, S.: Design of a time-sorting pitfall trap for surface-active arthropods. *Entomol. Exp. Appl.* 133, 100-103 (2009). <https://doi.org/10.1111/j.1570-7458.2009.00902.x>
14. McMunn, M.S.: A time-sorting pitfall trap and temperature datalogger for the sampling of surface-active arthropods. *HardwareX.* 1, 38-45 (2017). <https://doi.org/10.1016/j.ohx.2017.02.001>
15. Espinosa Z3niga, J.J.: Aplicaci3n de metodolog3a CRISP-DM para segmentaci3n geogr3fica de una base de datos p3blica. *Ing. Investig. y Tecnol.* 21, 1-13 (2020). <https://doi.org/10.22201/fi.25940732e.2020.21n1.008>

Estrategias Algorítmicas y Estructuras de Datos Eficientes para Búsquedas en Datos Masivos

Gabriel Tolosa, Santiago Bancho, Tomás Delvechio, Pablo Lavallén
Francisco Tonín Monzón, Agustín González, Esteban A. Ríssola
{tolosoft, sbancho, tdevechio, plavallen, ftonin, agonzalez, earissola}@unlu.edu.ar

Departamento de Ciencias Básicas, Universidad Nacional de Luján

Resumen

El mundo digital nos expone diariamente a una cantidad de datos constantemente creciente que exige contar con herramientas eficaces y muy eficientes para procesarlos y accederlos. La diversidad de aplicaciones que producen y consumen datos, sumada a un número también creciente de usuarios impone desafíos computacionales, tanto algorítmicos como del hardware disponible. Ejemplos típicos son sistemas de búsquedas de gran escala (como los motores de búsqueda web) o los servicios de búsqueda en tiempo real (como aquellos disponibles en las redes sociales).

Estos escenarios no solo exigen mayores capacidades a los proveedores de servicios (lo que impacta en su operación) sino, además, mejoras conceptuales y prácticas en las estructuras de datos y los algoritmos necesarios para que los sistemas escalen adecuadamente y puedan gestionar la demanda. La eficiencia es un requerimiento fundamental en el mundo digital actual caracterizado por datos masivos, heterogéneos y dinámicos.

Estas líneas de investigación abordan problemas de búsqueda en datos masivos, tanto desde las estructuras de datos como de los algoritmos necesarios para procesar documentos, publicaciones en redes sociales o consultas, con el objetivo de posibilitar la escalabilidad de los sistemas de búsqueda con el objetivo final de hacer un uso más racional de los recursos.

Palabras clave: algoritmos eficientes, búsquedas web, datos masivos.

Contexto

Esta presentación corresponde a los proyectos de investigación “Estrategias y Algoritmos para Problemas de Búsquedas a Gran Escala” (Disp.CDCB: 350/19) y “Algoritmos de Aprendizaje Automático para Respuestas en Tiempo Real sobre Entornos de Datos Masivos” (Disp.CDCB: 550/17), Departamento de Ciencias Básicas (UNLu).

Introducción

El mundo digital actual se caracteriza por la cantidad, variedad y velocidad a la que se producen y consumen datos y/o información: Estos atributos se los asocia, habitualmente, con el concepto de *Big Data* o datos masivos [20]. Una de las aplicaciones clave para el acceso eficaz y eficiente a la información en este escenario son los sistemas de búsqueda, tanto sobre colecciones de documentos, *posts* en redes sociales como Twitter [10, 25] (donde las publicaciones ocurren en tiempo real) o los grafos subyacentes que forman estos sistemas [15].

Además, una complejidad extra que se asocia es la necesidad de responder a los usuarios con restricciones de tiempo, en algunos casos muy estrictas. Por ejemplo, un motor de búsqueda de escala web como Google, que procesa miles de millones de documentos [7] y debe responder en una fracción de segundo (por ejemplo, 200 ms [31]). En un trabajo sobre el comportamiento de los usuarios, Arapakis y otros [3] reportaron que el incremento en la latencia reduce la posibilidad de *clicks* en resultados o avisos, lo que disminuye los retornos para los proveedores del servicio. Análogamente, Crescenzi y otros [12] encontraron que los usuarios en condiciones de restricción de tiempo debido al sistema informaron una presión mayor para completar sus tareas y una me-

nor satisfacción con su desempeño.

Luego, un requerimiento necesario es diseñar y evaluar estructuras de datos específicas y combinaciones de algoritmos sofisticados que permitan el procesamiento eficiente, considerando también parámetros de eficacia. Muchas de las técnicas diseñadas y aplicadas para este tipo de problemas son luego *exportadas* a otros problemas prácticos como las búsquedas en grafos [16], el procesamiento de secuencias biológicas (por ejemplo, genomas) [8] o los sistemas de recomendación [5]. Pero, de manera equivalente, técnicas de otros dominios de las Ciencias de la Computación son hoy en día muy útiles en el dominio de la Recuperación de Información como, por ejemplo, algoritmos de Aprendizaje Automático son aplicados al problema del ranking en motores de búsqueda [2] o el muestreo aplicado a grafos para estimar métricas globales [1].

En el caso específico de los motores de búsqueda web (MBW) se aplican técnicas de poda dinámica [29], caching [27], ranking [26] (como *Learning to Rank*), compresión de las estructuras de datos [24] o selección de recursos [18], entre otras. En general, estas aplicaciones se ejecutan en un cluster, por lo que las estrategias de distribución de la carga entre los nodos [32] y la partición de las estructuras de datos son de gran relevancia.

Procesar consultas en un MBW es uno de los desafíos más difíciles de manejar debido al constante crecimiento de datos y usuarios [22], donde el *tradeoff* entre la eficacia y la eficiencia es un factor clave [13].

Por ello, estas líneas de investigación abordan problemas relacionados con las estructuras de datos y algoritmos eficientes que permitan realizar búsquedas sobre datos masivos y escalen de acuerdo al tamaño del problema. Esto permite limitar el crecimiento de las infraestructuras de cómputo, disminuyendo los costos operativos y mejorando el impacto ambiental.

Líneas de I+D

En esta sección se presentan las líneas de I+D del grupo, las cuales se vienen desarrollando en varios proyectos de investigación en los últimos años.

a. Datos Masivos

a.1. Estructuras de Datos

Los sistemas de búsqueda utilizan como estructura de datos predominante un índice invertido el cual se compone, básicamente, por un vocabulario (V) con los términos (t_i) disponibles para la búsqueda. Luego, por cada $t \in V$ se asocia una lista de documentos (d_j) donde éstos aparecen (*posting lists*) junto con información asociada (por ejemplo, $f_{i,j}$, la frecuencia de t_i en d_j). Los índices invertidos son, en general, comprimidos para poder gestionarlos en memoria principal y acelerar su recorrido.

Compresión de Índices: El tamaño del índice invertido está dominado por el conjunto de *posting lists*, las cuales son identificadores de documentos con frecuencias asociadas. Por lo tanto, lograr reducir este tamaño resulta en un problema de compresión de números enteros (individualmente [24, 19, 35] o en listas [35, 19, 34, 30, 23]). Si bien hay muchos trabajos en los últimos años, el problema sigue aún siendo de interés en la comunidad ya que hay diferentes escenarios de uso y múltiples requerimientos. El objetivo final en cualquier caso es lograr un *tradeoff* entre espacio utilizado por el índice comprimido y tiempo de descompresión adecuado que permita cumplir con los requisitos de cada aplicación particular.

En esta línea se aborda la compresión de un índice invertido en bloques utilizando diferentes codecs (lo usual es comprimir una lista o todo el índice con un mismo algoritmo). La idea es particionar las listas de bloques con propiedades particulares que resulten más adecuadas para un determinado codec, resultando en un esquema *multicompresión* del cual ya se han desarrollado trabajos preliminares [17] y que continúan en una tesina de grado. En esta se considera, además, particiones de las listas en bloques de tamaño tanto uniforme como variable.

Índices Híbridos: Las *posting lists* de los índices invertidos se ordenan, habitualmente, de forma ascendente por identificador de documento (esto favorece la compresión). No obstante, es posible ordenarlas descendientemente por la frecuencia de los términos en los documentos (o factor de impacto), lo que permite hallar rápidamente “buenos” candidatos a participar de las listas de k resultados. Una posibilidad que se explora es una representación mixta

donde cada lista comience con un porcentaje i de documentos ordenados por frecuencia y el resto por número de documento. En esta línea se estudian las posibles combinaciones de codecs de compresión que permiten reducir el tiempo de procesamiento con un balance adecuado respecto del espacio utilizado.

a.2. Algoritmos

Las estrategias más eficientes para recorrer un índice invertido de colecciones de documentos masivos se basan en la idea de procesar un documento a la vez (DAAT, Document-at-a-Time) sobre listas ordenadas ascendentemente por el identificador de documento (d_j). Básicamente, a partir de un *query* q_i compuesto por n términos ($q_i = \{t_1, t_2 \dots t_n\}$), DAAT recorre las n listas en paralelo, computando el *score* de cada d_j . Las técnicas eficientes omiten aquellos documentos que no tienen posibilidad de formar parte de la lista final de k resultados (*dynamic pruning*). Entre éstas, las predominantes son Maxscore [29] y WAND [9] (con sus variantes basadas en bloques de longitud fija [14] o variable [21]). Estas técnicas se consideran *safe* ya que retornan exactamente el mismo resultado como si se hubiesen evaluado todos los documentos.

Top-k: Las técnicas de *dynamic pruning* se basan en aprovechar un valor umbral (*upper bound*) asociado a cada lista (o bloque) que permita determinar en qué momento finalizar la evaluación. La propuesta de esta línea es aumentar las capacidades de *dynamic pruning* de Maxscore usando múltiples valores umbrales para una misma lista. Se construye una estructura similar a las *skip list* [11] que no impone un *overhead* significativo y que se carga en memoria antes de resolver la consulta. Los resultados preliminares son muy alentadores, posibilitando hasta un 2x de *speedup* en la evaluación.

La segunda propuesta se basa en la representación híbrida del índice (presentada en la sección anterior) junto con un algoritmo que explota esta característica. Cada sección se recorre usando una técnica apropiada y la idea es contar con los documentos candidatos más rápidamente para que la poda dinámica resulte más eficiente. Dado que esta representación no favorece la compresión del índice, un desafío es hallar un *tradeoff* aceptable entre el tamaño de la primera sección y la reducción en los tiempos de procesamiento que ofrece. Por otro lado, si el recorrido de

las listas no es exhaustivo, la estrategia no resulta *safe* por lo que se estudia el impacto de esta idea.

Caching: Otra de las técnicas algorítmicas que se investigan son diversas políticas de *caching* asociadas al problema de las búsquedas. En un motor de búsquedas se pueden utilizar diferentes niveles de caches (por ejemplo, resultados, posting lists, intersecciones). Los últimos avances son las propuestas de Yafay y otros [33] orientadas a mantener en memoria los scores de la k -ésima posición de la lista de resultados de diferentes consultas.

La propuesta en esta línea de investigación es poder mantener en un cache los identificadores de documentos más pequeño y más grande que participan de una lista de resultados para un query y un valor k determinados (la técnica se denomina “MinMaxCaching”). Esto permite acelerar el procesamiento de los siguientes queries permitiendo avanzar los cursores de las listas invertidas de los términos en q hasta el valor mínimo en cache y luego finalizar la evaluación una vez alcanzado el valor máximo. Los experimentos iniciales muestran que se puede acelerar Maxscore hasta un 30% para algunas consultas complejas. Se está evaluando, además, aplicar la propuesta a BlockMax-Wand [14].

Flujos en Tiempo Real: Las búsquedas sobre flujos que ocurren en tiempo real requieren de algoritmos y estructuras de datos dinámicas que escalen a, por ejemplo, millones de usuarios en una red social como Twitter. El desafío es ofrecer el contenido en un período de tiempo extremadamente corto luego de cada publicación. Este problema puede ser abordado particionando la colección en porciones (*shards*) que se distribuyen en p nodos. Luego, una búsqueda particular se envía solo a un número reducido de nodos (los que se estiman como los más *adecuados* para resolverla). Este enfoque se denomina “búsquedas selectivas” (*selective search*).

En esta línea de investigación se proponen diferentes criterios de actualización del índice invertido y su particionado, como así también nuevas estrategias de selección de recursos para el algoritmo de búsqueda. Dado que una arquitectura para este problema requiere de un sistema distribuido ejecutándose en un cluster, el balance de carga es un factor importante a considerar en las propuestas.

Finalmente, una línea que combina caching con

flujos en tiempo real consiste en el desarrollo de políticas de admisión para un cache de resultados basada en algoritmos para *streaming*. Concretamente, se aplican clasificadores que *aprenden* patrones de los queries utilizando Hoeffding Adaptive Trees [6] para detectar *singletons* (aquellos que aparecen solo una vez). Dado que las distribuciones de aparición de consultas son muy sesgadas (siguen una ley de potencias [4]) y aproximadamente un 50% de consultas son únicas, detectarlas se vuelve importante para evitar ingresarlas al cache y desplazar otra con potencial de aparecer nuevamente. Algunos resultados preliminares [28] sugieren que esta aproximación es válida y puede ser utilizada eficientemente para aumentar la performance de un cache de resultados en un motor de búsquedas.

Objetivos y Resultados

Las líneas de investigación mencionadas abordan problemas de búsquedas en datos masivos, las cuales representan desafíos [13] y oportunidades de transferencia. Por ello, los objetivos principales son:

- Diseñar y evaluar variantes de los algoritmos para recuperación *top-k* haciendo usos alternativos de los *upper bounds* de las listas y usando umbrales de poda/corte dinámicos para reducir el número de elementos procesados.
- Desarrollar estructuras de datos (y algoritmos de recorrido) que combinen ordenamientos mixtos de las listas basados en frecuencias e identificadores de documentos, de forma variable para optimizar la poda dinámica.
- Desarrollar y evaluar esquemas multicompresión de listas que combinen codecs apropiados a bloques de datos con propiedades específicas y consideren, además, parámetros del problema donde se apliquen (por ejemplo, mayor necesidad de compresión).
- Diseñar y evaluar estrategias de indexación distribuida (y políticas de asignación de documentos) y resolución de consultas para flujos en tiempo real. Esto incluye métodos de selección de nodos y técnicas de *caching*.
- Definir y evaluar estructuras y modelos de cómputo distribuido sobre clusters de *hardware commodity* para problemas de escalabilidad en ambientes de búsquedas selectivas.

- Definir estrategias de caching para identificadores de documentos (MinMaxCaching) y políticas de admisión para caches de resultados usando algoritmos de aprendizaje sobre streams de queries.

Formación de Recursos Humanos

En el marco de estas líneas de investigación se están dirigiendo tres tesis de Licenciatura en Sistemas de Información (UNLu). Además, asociados a los proyectos de investigación mencionados hay una Beca Estímulo a las Vocaciones Científicas (CIN) y dos pasantías internas en la UNLu.

Referencias

- [1] N. K. Ahmed, N. Duffield, T. L. Willke, and R. A. Rossi. On sampling from massive graph streams. *Proc. VLDB Endow.*, 10, 2017.
- [2] Q. Ai, T. Yang, H. Wang, and J. Mao. Unbiased learning to rank: Online or offline? *ACM Trans. Inf. Syst.*, 39(2), Feb. 2021.
- [3] I. Arapakis, X. Bai, and B. B. Cambazoglu. Impact of response latency on user behavior in web search. SIGIR '14, page 103–112, New York, NY, USA, 2014. ACM.
- [4] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval: The Concepts and Technology Behind Search*. Addison-Wesley Publishing Company, 2nd edition, 2011.
- [5] A. Bellogin and A. Said. *Information Retrieval and Recommender Systems*, pages 79–96. Springer International Publishing, Cham, 2019.
- [6] A. Bifet and R. Gavaldà. Adaptive learning from evolving data streams. In N. Adams, C. Robardet, A. Siebes, and J. Boulicaut, editors, *Advances in Intelligent Data Analysis VIII*, 2009.
- [7] A. Bosch, T. Bogers, and M. Kunder. Estimating search engine index size variability: a 9-year longitudinal study. *Scientometrics*, 107(2):839–856, May 2016.
- [8] C. Boucher. Indexing genomes in a scalable manner (keynote). In *28th Int. Symposium of String Processing and Information Retrieval*, SPIRE'21, 2003.
- [9] A. Z. Broder, D. Carmel, M. Herscovici, A. Soffer, and J. Zien. Efficient query evaluation using a two-level retrieval process. In *Proc. of the 12th Int. Conf. on Information and Knowledge Management, CIKM '03*, 2003.

- [10] M. Busch, K. Gade, B. Larson, P. Lok, S. Luckenbill, and J. Lin. Earlybird: Real-time search at twitter. In *28th Int. Conf. on Data Engineering, ICDE '12*. IEEE, 2012.
- [11] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to algorithms*. MIT press, 2009.
- [12] A. Crescenzi, D. Kelly, and L. Azzopardi. Impacts of time constraints and system delays on user experience. In *In 2016 ACM on Conference on Human Information Interaction and Retrieval, CHIIR '16*, page 141–150. ACM, 2016.
- [13] J. S. Culpepper, F. Diaz, and M. D. Smucker. Research frontiers in information retrieval (swirl 2018). *SIGIR Forum*, 52(1), 2018.
- [14] S. Ding and T. Suel. Faster top-k document retrieval using block-max indexes. In *Proc. of the 34th Int. Conf. on Research and Development in Information Retrieval*. ACM, 2011.
- [15] M. Douze, A. Sablayrolles, and H. Jégou. Link and code: Fast indexing with s and compact regression codes. In *The IEEE Conf. on Computer Vision and Pattern Recognition*, 2018.
- [16] A. Giordano and G. Tolosa. Improved landmark-based shortest path length estimation in large graphs with distance correction. In *IEEE/WIC/ACM Int. Joint Conf. on Web Intelligence and Intelligent Agent Tech.*, 2020.
- [17] A. González and G. Tolosa. Multicompresión de grandes listas de enteros para sistemas de búsquedas. In *Simposio Argentino de GRANdes DATos*. JAIIO, 2020.
- [18] Y. Kim, J. Callan, J. S. Culpepper, and A. Moffat. Efficient distributed selective search. *Information Retrieval Journal*, 20(3), June 2017.
- [19] D. Lemire and L. Boytsov. Decoding billions of integers per second through vectorization. *Softw. Pract. Exper.*, 45(1), Jan. 2015.
- [20] S. Madden. From databases to big data. *IEEE Internet Computing*, 16(3), 2012.
- [21] A. Mallia, G. Ottaviano, E. Porciani, N. Tonellotto, and R. Venturini. Faster blockmax wand with variable-sized blocks. In *Proc. of the 40th Int. Conf. on Research and Development in Information Retrieval*, 2017.
- [22] A. Mallia and E. Porciani. Faster blockmax wand with longer skipping. In L. Azzopardi, B. Stein, N. Fuhr, P. Mayr, C. Hauff, and D. Hiemstra, editors, *Advances in Information Retrieval*. Int. Publishing, 2019.
- [23] G. Ottaviano and R. Venturini. Partitioned elias-fano indexes. In *37th Int. ACM SIGIR Conf. on Research & Development in Information Retrieval, SIGIR '14*, 2014.
- [24] G. E. Pibiri and R. Venturini. Techniques for inverted index compression. *ACM Comput. Surv.*, 53(6), 2020.
- [25] E. Rissola and G. Tolosa. Improving real time search performance using inverted index entries invalidation strategies. *Journal of Computer Science & Technology*, 16(1), 2016.
- [26] N. Tax, S. Bockting, and D. Hiemstra. A cross-benchmark comparison of 87 learning to rank methods. *Information Processing & Management*, 51, 2015.
- [27] G. Tolosa, E. Feuerstein, L. Becchetti, and A. Marchetti-Spaccamela. Performance improvements for search systems using an integrated cache of lists + intersections. *Information Retrieval. Journal*, 20(3):172–198, 2017.
- [28] F. Tonin-Monzón, S. Banchemo, and G. Tolosa. Árboles de decisión adaptativos en políticas de admisión a caché. In *Simposio Argentino de GRANdes DATos*. JAIIO, 2018.
- [29] H. Turtle and J. Flood. Query evaluation: Strategies and optimizations. *Inf. Process. Manage.*, 31(6), Nov. 1995.
- [30] S. Vigna. Quasi-succinct indices. In *Sixth ACM Int. Conf. on Web Search and Data Mining, WSDM '13*, 2013.
- [31] L. Wang, D. Metzler, and J. Lin. Ranking under temporal constraints. In *Proceedings of the 19th ACM Int. Conf. on Information and Knowledge Management, CIKM '10*, page 79–88. ACM, 2010.
- [32] Y. Wang, L. Wu, L. Luo, Y. Zhang, and G. Dong. Short-term internet search using makes people rely on search engines when facing unknown issues. *PLoS One*, 12(4), 2017.
- [33] E. Yafay and I. S. Altingovde. Caching scores for faster query processing with dynamic pruning in search engines. In *In 28th ACM Int. Conf. on Information and Knowledge Management*, page 2457–2460. ACM, 2019.
- [34] H. Yan, S. Ding, and T. Suel. Inverted index compression and query processing with optimized document ordering. In *18th Int. Conf. on World Wide Web*, 2009.
- [35] J. Zhang, X. Long, and T. Suel. Performance of compressed inverted list caching in search engines. In *17th Int. Conf. on World Wide Web*, 2008.

Hacia la definición de un marco metodológico para el desarrollo de un sistema de reconocimiento biométrico mediante técnicas de Machine Learning

Silvia E. Ruiz, Carlos E. Alvez, Graciela R. Etchart, Ernesto M. Miranda
Facultad de Cs. de la Administración - UNER
{silvia.ruiz, carlos.alvez, graciela.etchart, ernesto.miranda}@uner.edu.ar

Resumen

En el mundo moderno de hoy, con el avance de la ciencia y la tecnología, el reconocimiento de las personas se ha convertido en una necesidad. La biometría es la rama de la ciencia que se ocupa de los métodos automatizados para reconocer una persona en función de una característica fisiológica o conductual tales como rostro, huellas digitales, iris, retina, firma, voz, entre otras. Dentro de estos rasgos, el iris se destaca por presentar una estructura única e invariable durante toda la vida del ser humano. Con el objetivo de mejorar las etapas que conforman un sistema de reconocimiento de iris, se encuentran en auge el uso de distintas técnicas de machine learning (ML) [1,2]. En este sentido, el uso de una metodología en el proceso de desarrollo de un sistema permite obtener productos más confiables [3] y constituye una guía para planificar y estructurar el proyecto. La línea de investigación aquí presentada propone la definición de un marco metodológico adecuado a las particularidades de un sistema de reconocimiento biométrico de iris mediante técnicas de machine learning.

Palabras clave: sistema biométrico, metodologías, machine learning.

Contexto

La presente línea de investigación forma parte del desarrollo de un trabajo de tesis de la Maestría en Sistemas de Información de la Facultad de Ciencias de la Administración perteneciente a la Universidad Nacional de Entre Ríos.

Además, este trabajo da continuidad a las líneas de investigación iniciadas dentro del marco de trabajo del Proyecto PID 07/G058 “*Modelos de Machine Learning para la mejora de la precisión, seguridad y*

eficiencia en la gestión de datos biométricos” y sus antecesores.

Introducción

Se observa en la actualidad una tendencia muy marcada en el empleo de rasgos biométricos, tales como rostro, huellas digitales, geometría de la mano, iris, patrones de retina, firma y voz, entre las más destacadas en los sistemas de identificación/autenticación de personas. Todas estas técnicas presentan diferentes grados de singularidad, permanencia, mensurabilidad, desempeño, aceptación del usuario y robustez.

Los sistemas de reconocimiento biométrico usan características fisiológicas o de comportamiento propias de cada individuo para identificarlo, es decir, se reconoce al usuario por lo que es en lugar de por lo que tiene o sabe [4]. Los rasgos fisiológicos presentan una reducida variabilidad a lo largo del tiempo, pese a que su adquisición es más invasiva y requiere de la cooperación de los sujetos. Por el contrario, los rasgos de comportamiento resultan menos invasivos aunque la exactitud de la identificación es menor debido a la variabilidad de los patrones de comportamiento.

Dentro de los rasgos fisiológicos, la textura del iris proporciona un alto grado de singularidad y estabilidad, por lo que es muy poco probable que dos patrones de iris sean iguales. Debido a esto, se lo considera uno de los rasgos biométricos más seguros contra métodos fraudulentos y ataques de suplantación de identidad [5].

El reconocimiento de iris se encuentra en constante crecimiento y es considerado un área de investigación importante y desafiante, a pesar de ser un campo de estudio muy reciente en relación a otros rasgos biométricos [6]. En

este sentido, a lo largo de los últimos años, se han comenzado a utilizar diferentes técnicas de ML en las etapas que conforman el proceso de reconocimiento de iris [1]. En [1,2] se destacan principalmente las técnicas de ML tales como SVM (Máquina de Soporte Vectorial), WPNN (Redes Neuronales Probabilísticas Ponderadas), BPNN (Redes Neuronales de Retro Propagación) y RBFNN (Redes Neuronales con Función de Base Radial), las cuales son utilizadas generalmente en la etapa de clasificación de los patrones del iris, así como también para la normalización, segmentación y extracción de características.

Por otro lado, en toda implementación de un sistema es conveniente seguir un modelo de proceso o metodología [7] a fin de obtener mayor probabilidad de éxito en los resultados obtenidos. Según Pressman [7], un modelo de proceso define un conjunto de actividades organizadas para llevar a cabo una tarea determinada, mientras que una metodología además de definir las actividades y tareas de un proceso, define cómo llevar a cabo las mismas.

Al respecto, si bien en la actualidad se están desarrollando distintas líneas de investigación para definir una metodología estándar específica para utilizar en sistemas que utilicen ML [8] no se ha encontrado en la literatura revisada una definición estándar formal. Existen varias metodologías y modelos de procesos de data mining y ML, tales como KDD [9] (**K**nowledge **D**iscovery in **D**atabases) y SEMMA (**S**ample, **E**xplore, **M**odify, **M**odel and **A**ssess) desarrollada por el Instituto SAS (Statistical Analysis Systems)¹, sin embargo se suele adoptar como base la metodología estándar CRISP-DM (**C**ross **I**ndustry **S**tandard **P**rocess for **D**ata **M**ining) [10]. Este estándar es utilizado habitualmente como marco metodológico en distintos tipos de proyectos [11], principalmente de data mining.

Es por ello que, en el presente trabajo se presenta como línea de investigación la definición de un marco metodológico para

implementar un sistema biométrico de iris mediante técnicas de ML.

Líneas de investigación y desarrollo

Los principales ejes temáticos de la presente investigación son:

Sistemas biométricos

Un sistema biométrico es esencialmente un sistema de reconocimiento de patrones que opera adquiriendo datos biométricos de un individuo, extrayendo un conjunto de características de los datos adquiridos y comparando este conjunto de características con la plantilla almacenada en la base de datos. En la Figura 1 se observa un diagrama de bloque representativo de un sistema biométrico.

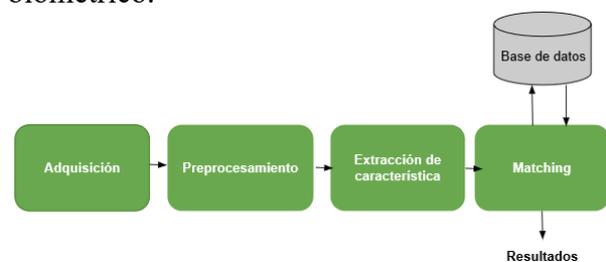


Figura 1: Diagrama de bloque de un sistema biométrico

Un sistema biométrico permite reconocer a una persona en base a un vector de características fisiológicas o conductuales específicas que posee la persona. Sin embargo, para que un sistema biométrico sea útil se deben considerar cuestiones de aceptabilidad y rendimiento. Es decir, el sistema debe ser aceptado por la población destinataria y ser lo suficientemente robusto frente a distintos métodos y ataques fraudulentos.

Machine Learning

El aprendizaje automático es un campo de la informática que estudia algoritmos y técnicas para automatizar soluciones a problemas complejos que son difíciles de programar utilizando métodos de programación convencionales [12].

En la actualidad, se utilizan una variedad de algoritmos que aprenden iterativamente de los datos para mejorar, describir datos y predecir resultados. A medida que los algoritmos utilizan datos de entrenamiento, es posible producir modelos más precisos basados en

¹ https://www.sas.com/es_ar/home.html

esos datos. Un modelo de machine learning es el resultado generado cuando se entrena un algoritmo de aprendizaje automático con datos. Esto representa una de las diferencias fundamentales con una metodología tradicional de desarrollo de software, ya que la utilización de ML involucra al menos un modelo ML que utiliza datos de entrenamiento y procesa datos adicionales para hacer algunas inferencias. En este sentido, en [13] se propone un proceso para resolver problemas de ciencias de datos denominado “Hoja de ruta de la ciencia de datos” (Figura 2).

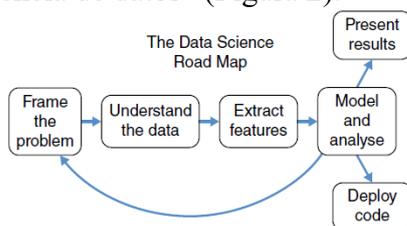


Figura 2: Hoja de ruta de la ciencia de datos obtenida de [13].

Metodologías

El modelo KDD propuesto en [9], se compone por cinco fases: selección, preprocesamiento, transformación, minería de datos y evaluación e implantación. En general, los términos de descubrimiento de conocimiento en base de datos y data mining, suelen utilizarse de manera indistinta [14], aunque existen diferencias entre ambos. El término KDD hace referencia a un proceso que consta de una serie de fases, mientras que data mining es sólo una de estas fases.

Con el paso del tiempo fueron surgiendo nuevas metodologías tales como SEMMA y CRISP-DM. Ésta última constituye un marco metodológico de referencia y se encuentra más orientado a la analítica y la obtención de modelos basados en datos. Tal como lo describe Miralles en [15] las fases y actividades propuestas por esta metodología constituyen una buena guía para el proceso de descubrimiento de conocimientos.

A pesar de existir distintas metodologías, resulta importante destacar que para que una metodología o modelo de proceso resulte útil como guía de trabajo, es necesario adaptarlo a las necesidades específicas de un proyecto. En este sentido, teniendo en cuenta las particularidades que posee un sistema de reconocimiento biométrico de iris mediante

técnicas de machine learning, en el presente trabajo se proponen las siguientes fases (Figura 3):

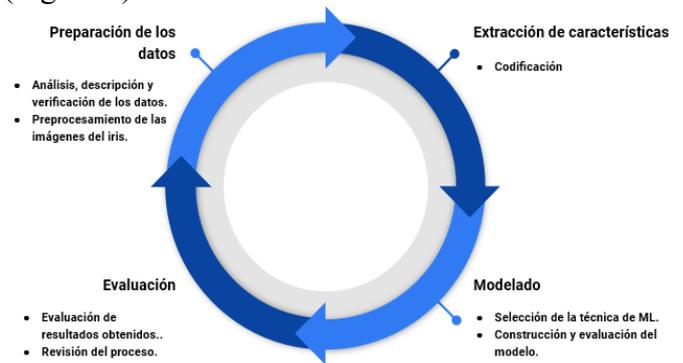


Figura 3: Fases de la metodología propuesta.

A continuación se describen brevemente las fases del esquema propuesto:

- *Preparación de los datos:* Implica las tareas de análisis, descripción, verificación y preprocesamiento de los datos. En cuanto a la tarea de descripción de los datos, es necesario establecer volúmenes de datos, su identificación, el significado de cada campo y la descripción del formato inicial. Cabe aclarar, que siempre es necesario efectuar verificaciones para asegurar la completitud y corrección de los datos, de manera de determinar la consistencia de los valores individuales de los campos, la cantidad y distribución de los valores nulos, y para encontrar valores fuera de rango, los cuales pueden constituirse en ruido para el proceso. Antes de extraer las características de la imagen original del ojo es necesario preprocesar la imagen para localizar los límites de la pupila y el iris y así poder normalizar la región de interés. En [16] se demuestra que un sistema de reconocimiento de iris mediante un enfoque de aprendizaje profundo (deep learning) resulta más apropiado que los enfoques tradicionales para mejorar las técnicas de segmentación y detectar imágenes de iris falsas.
- *Extracción de características:* en esta fase se procede a la extracción de características y su correspondiente codificación. Para ello, pueden ser utilizados métodos tradicionales tales

como codificación Wavelet, filtro Gabor, filtro Log-Gabor, filtro de Gabor 2D, entre otros, así como también técnicas de ML [17]. Es importante destacar que, el éxito de cualquier sistema biométrico definido como un sistema de clasificación y reconocimiento depende principalmente de la eficiencia y robustez de las etapas de extracción y clasificación de características.

- *Modelado*: esta fase abarca el diseño, construcción y evaluación del modelo. Aquí es necesario seleccionar las técnicas de modelado más apropiadas para la construcción del modelo. En general, existen varias técnicas para el mismo tipo de problema de machine learning. Algunas tienen requisitos específicos sobre la forma de los datos. Debido a esto, muchas veces resulta necesario volver a la fase de preparación de los datos. Los parámetros utilizados en la generación del modelo, dependen de las características de los datos y de la precisión que se desea lograr con el modelo. En general, las técnicas más utilizadas como clasificadores son SVM y redes neuronales [2]. Una vez construido un modelo, se debe generar un procedimiento destinado a probar la calidad y validez del mismo. En general, se separan los datos en dos conjuntos, uno de entrenamiento y otro de prueba, para luego construir el modelo basado en el conjunto de entrenamiento y medir la calidad del modelo generado con el conjunto de prueba. Después de seleccionada la técnica, se ejecuta sobre los datos previamente preparados para generar uno o más modelos. La selección de los mejores parámetros es un proceso iterativo y se basa exclusivamente en los resultados generados. Estos deben ser interpretados y su rendimiento justificado.
- *Evaluación*: en esta fase se evalúan los resultados obtenidos. Es importante evaluar a fondo y revisar los pasos ejecutados en la creación del modelo de manera de asegurar que este logre adecuadamente los objetivos. Para ello, se

realizan diferentes entrenamientos con distintos algoritmos y parámetros de los mismos.

Resultados obtenidos/esperados

Objetivo general

Proponer un marco metodológico adaptado a las necesidades de un sistema de reconocimiento biométrico mediante técnicas de ML.

Objetivos específicos

1. Realizar un relevamiento bibliográfico relacionado a los conceptos fundamentales de la biometría y los sistemas biométricos.

2. Explorar y analizar los fundamentos teóricos, técnicas y herramientas de los distintos algoritmos de ML en base a sus características y potencialidades.

3. Analizar el uso de los distintos mecanismos de ML aplicados actualmente en los sistemas de reconocimiento biométrico.

4. Realizar un estudio de las principales metodologías y/o modelos de procesos utilizados actualmente en proyectos de ML y específicamente en sistemas biométricos.

5. Definir los criterios de evaluación que permitan realizar un análisis comparativo de las distintas metodologías y/o procesos.

6. Proponer un marco metodológico que constituya una guía en base a las distintas fases, subfases y actividades para implementar en un sistema biométrico de reconocimiento de iris utilizando técnicas de ML.

Con respecto al cumplimiento de los objetivos propuestos, a la fecha se avanzó en las actividades relacionadas al análisis de las distintas fuentes bibliográficas correspondientes a los ejes temáticos de los sistemas biométricos, las técnicas de ML y en las distintas metodologías empleadas en proyectos de ML y data mining. También se avanzó en las tareas relacionadas a la primera fase propuesta en el marco metodológico "*Preparación de los Datos*". Para ello, se exploraron distintos conjuntos de datos (dataset) de libre acceso y se analizaron las ventajas y desventajas que presentan cada uno: CASIA, UBIRIS, entre otros.

En la Figura 4 se presenta una tabla con las

principales características de las bases de datos evaluadas:

Dataset	Tipo	Individuos	Capturas/ Individuos	Resolución	Espectro
UBIRIS V1	Ideal	241	7 (±)	2560 x 1704	VW
UBIRIS V2	No restringida	261	40 (±)	400 x 300	VW
UPOL	Ideal – menos restring.	64	6	576 x 768	VW
MMU	-	45	10	baja	-
IITD	Ideal – menos restring.	224	10	320 x 240	NIR
ICE-2005	-	124	12	480 x 640	NIR
Bath UBBID	Ideal	400	40	960 x 1280	NIR

Figura 4: Tabla comparativa de bases de datos de iris.

Además se avanzó en la propuesta de un marco metodológico con cuatro fases principales. No obstante, aún queda mucho trabajo por hacer profundizando las distintas actividades que componen cada una de las fases propuestas en el marco metodológico, y en especial en determinar la opción más apropiada de las técnicas de ML a utilizar para la construcción del modelo. También resta explorar y analizar las distintas métricas de ML de manera tal de poder evaluar el proceso realizado.

Formación de recursos humanos

En la presente línea de investigación se enmarca el desarrollo de una tesis de maestría. Además, en el proyecto de investigación mencionado en la sección “contexto” se cuenta con la participación de un becario de Iniciación en la Investigación.

Referencias

- [1] M. De Marsico, A. Petrosino, y S. Ricciardi, “Iris recognition through machine learning techniques: A survey”, *Pattern Recognit. Lett.*, vol. 82, pp. 106-115, oct. 2016.
- [2] A. A. Khan, S. Kumar, y M. Khan, “Iris Pattern Recognition using Support Vector Machines and Artificial Neural Networks”, *IJIREEICE*. 2014.
- [3] Ghezzi C., Jazayeri M., Mandrioli D., “*Fundamentals of Software Engineering*”, 2nd. Edition. Pearson, 2002.
- [4] A. K. Jain, A. Ross, y S. Prabhakar, “An introduction to biometric recognition”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, n.º 1, pp. 4-20, ene. 2004.
- [5] B. Arslan, E. Yorulmaz, B. Akca, y S. Sagioglu, “Security Perspective of Biometric Recognition and Machine Learning Techniques”, en 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), dic. 2016, pp. 492-497.
- [6] J. G. Daugman, “High confidence visual recognition of persons by a test of statistical independence”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, n.º 11, pp. 1148-1161, nov. 1993.
- [7] R. S. Pressman y B. R. Maxim, *Software engineering: a practitioner’s approach*, Ninth edition. New York, NY: McGraw-Hill Education, 2020.
- [8] S. Studer *et al.*, «Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology», *Cs Stat*, mar. 2020.
- [9] U. Fayyad, G. Piatetsky-Shapiro, y P. Smyth, «*The KDD process for extracting useful knowledge from volumes of data*», *Commun. ACM*, vol. 39, n.º 11, pp. 27-34, nov. 1996.
- [10] C. Shearer, «*The CRISP-DM model: the new blueprint for data mining*», *J Data Wareh.*, vol. 5, pp. 13-22, ene. 2000.
- [11] V. G. Cortina, “*Aplicación de la Metodología CRISP-DM a un Proyecto de Minería de Datos en el Entorno Universitario*”, p. 120.
- [12] G. Rebalá, A. Ravi, S. Churiwala. “*An Introduction to Machine Learning*”. Springer. 2019.
- [13] F. Cady. “*The Data Science Handbook*”. Wiley. 2017.
- [14] J. Hernández Orallo, C. Ferri Ramírez, y M. J. Ramírez Quintana, “*Introducción a la minería de datos*”. Madrid: Pearson Educación SA, 2004.
- [15] J. Miralles Solé. “*Proyectos de Inteligencia Artificial*”. 2020.
- [16] J. J. Winston and D. J. Hemanth, “A comprehensive review on iris image-based biometric system” *Soft Computing*, vol. 23, no. 19, pp. 9361–9384, 2019.
- [17] S. Minaee, A. Abdolrashidi, and Y. Wang, “An experimental study of deep convolutional features for Iris recognition shervin”, SPMB, Philadelphia, USA, 2016.

Determinación de perfiles socioeconómicos y sanitarios de las personas atendidas en las campañas efectuadas como actividades curriculares de la carrera de Medicina de la UNCAus en su área de influencia

David L. La Red Martínez, Stella Maris Gerzel, Emanuel Aguirre
Rubén A. Cáceres, José A. Berecochea Galarza, Selene Vallejo

Universidad Nacional del Chaco Austral
Sáenz Peña, (3700), Argentina

laredmartinezdavid@uncaus.edu.ar, stellagerzel@uncaus.edu.ar, aguirremanuel@uncaus.edu.ar,
rubens.uncaus@gmail.com, joseberecochea@gmail.com, selenevfores@gmail.com

Resumen

El proyecto PI N° 91, aprobado por Res. N° 332/18 CS, se desarrolla con el fin de obtener información relevante que detecte variables relacionadas con los frecuentes problemas de salud de la ciudad de Presidencia Roque Sáenz Peña (Chaco, Argentina) y su área de influencia, relacionando los pacientes con su hábitat, ecología y salud. Para ello se trabaja con la información proveniente de las actividades curriculares de vinculación comunitaria que la carrera de Medicina de la UNCAus (Universidad Nacional del Chaco Austral) realiza en los distintos barrios de la ciudad y zona de influencia; con dicha información se construye un almacén de datos (data warehouse) que es estudiado con técnicas de minería de datos, especialmente técnicas de agrupamiento (clusterización) y de árboles de decisión, a los efectos de conseguir los perfiles característicos relacionados con los distintos tipos de diagnósticos; se buscan modelos descriptivos y predictivos de minería de datos, lo cual permitiría disponer de conocimiento que permitiría mejorar la toma de decisiones en cuanto a campañas de salud hacia la población de las zonas.

Palabras Clave: minería de datos, soporte de decisión, perfiles socioeconómicos, perfiles

sanitarios, minería descriptiva, minería predictiva, salud humana.

Contexto

El desarrollo del proyecto responde a la necesidad de estructurar y generar información, a partir de datos obtenidos de las campañas de Actividades de Vinculación Comunitaria que realiza la carrera de Medicina como actividades curriculares, las cuales se llevan a cabo en los diferentes barrios y zonas de influencia de la localidad de Presidencia Roque Sáenz Peña, Argentina.

Se cuenta con gran cantidad de datos referidos a la condición socio económica y de salud de los pacientes, pero estos datos son representados como fichas médicas en papeles. En este sentido resulta importante digitalizarlos para luego aplicar técnicas de almacenes de datos, análisis multidimensional y minería de datos descriptiva y a posteriori predictiva que permitan obtener nuevos conocimientos. Los conocimientos generados a partir de los resultados de esta investigación permitirán encontrar y estudiar los perfiles socioeconómicos y sanitarios de los distintos barrios de la población y sus relaciones, elaborar propuestas de intervención y fundamentar acciones de mejora.

Introducción

La Minería de Datos (del inglés Data Mining - DM), es un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias en grandes volúmenes de datos, utilizando técnicas de AI y ML. Estas técnicas permiten extraer patrones y tendencias para describir y comprender mejor los datos y predecir comportamientos futuros.

La DM es la etapa de descubrimiento en el proceso de KDD (Knowledge Discovery from Databases), es el paso consistente en el uso de algoritmos concretos que generan una enumeración de patrones a partir de los datos preprocesados [1] [2] [3].

La DM es un conjunto de técnicas de análisis de datos que permiten extraer patrones, tendencias y regularidades para describir y comprender mejor los datos y extraer patrones y tendencias para predecir comportamientos futuros [4] [5] [6] [7].

Las técnicas de DM son diversas, una de las más utilizadas es la de clustering (o agrupamiento de datos) que consiste en la partición de un conjunto de individuos en subconjuntos lo más homogéneos posible, el objetivo es maximizar la similitud de los individuos del cluster y maximizar la diferencia entre clusters. El cluster demográfico es un algoritmo desarrollado por IBM, que resuelve automáticamente los problemas de definición de métricas de distancia / similitud, proporcionando criterios para definir una segmentación óptima [8] [9] [10] [11].

La DM es también un mecanismo de explotación, consistente en la búsqueda de información valiosa en grandes volúmenes de datos. Está muy ligada a los DW ya que los mismos proporcionan la información histórica con la cual los algoritmos de minería obtienen la información necesaria para la toma de decisiones [12] [13].

Un DW es una colección de datos orientada a un dominio, integrada, no volátil y variante en el tiempo para ayudar a tomar decisiones [14].

Un DW es una colección de datos orientado a temas, integrado, no volátil, de tiempo variante, que se usa para el soporte del proceso de toma de decisiones gerenciales.

Un DW es también un conjunto de datos integrados orientados a una materia, que varían con el tiempo, y que no son transitorios, los cuales soportan el proceso de toma de decisiones de una administración [15] [16] [4] [17].

Las tecnologías de DW y DM se vienen utilizando ampliamente en el ámbito de la salud. A continuación, se referencian algunos trabajos al respecto: [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31] [32] [33] [34] [35] [36], entre muchos otros.

Líneas de Investigación y Desarrollo

Determinar, mediante las técnicas de minería de datos, los factores o variables que causan las enfermedades crónicas frecuentes de las personas, relacionando el paciente con su hábitat, ecología y salud; utilizando los determinantes sociales de la salud como variables.

Utilizar un sistema desarrollado para ingresar la información en una base de datos, luego establecer las variables de estudio y aplicar las técnicas de minería de datos. Analizar los resultados obtenidos y confeccionar informes.

Resultado Esperados

Lograr explicitar en qué medida los determinantes sociales de la salud en los distintos barrios influyen en su situación sanitaria, utilizando variables relacionadas con el hábitat, la ecología y la salud, considerando como indicadores de esta las enfermedades crónicas que poseen los habitantes.

Se consideran las siguientes variables:

- Biología humana: edad, género, antecedentes biológicos y enfermedades crónicas no transmisibles.
- Medio ambiente: vivienda, agua y excretas.

- Socioeconómico: trabajo, cobertura de salud y nivel educacional alcanzado.
- Sistema de salud: embarazos, inmunizaciones y centros de salud cercanos.
- Estilo de vida: tabaquismo.
- Educación: máximo nivel educativo alcanzado.

En base a estas variables, y utilizando técnicas de clusterización, clasificación y determinación de patrones se espera establecer perfiles y situaciones problemáticas potenciales.

Los resultados logrados hasta el momento se han publicado en [37] y [38].

Formación de Recursos Humanos

El equipo de trabajo está compuesto por un Doctor, una Licenciada en Sistemas de Información, dos Ingenieros en Sistemas, un Profesor en Ciencias Químicas y del Ambiente y una becaria estudiante avanzada de Medicina.

Referencias

- [1] U. Fayyad, G. Grinstein, A. Wierse. **Information Visualization in Data Mining and Knowledge Discovery**. San Francisco, USA: Morgan Kaufmann Publishers. 2001.
- [2] U. Fayyad, G. Piatetskiy-Shapiro, P. Smith, U. Ramasamy. **Advances in Knowledge Discovery and Data Mining**. USA: AAAI Press / MIT Press. 1996.
- [3] D. Hand, H. Mannila, P. Smyth. **Principles of Data Mining**. USA: The MIT Press. USA. 2000.
- [4] A. Simon. **Data Warehouse, Data Mining and OLAP**. USA: John Wiley & Sons. 1997.
- [5] A. Berson, S. J. Smith. **Data warehouse, data mining & OLAP**. USA. Mc Graw Hill. 1997.
- [6] W. Frawley, G. Piatetsky-Shapiro, Ch. Matheus. **Knowledge Discovery in Database An Overview**. En Piatetsky-Shapiro, G Frawley, W. J (Ed) Knowledge discovery in databases (pp. 1-31). USA: The AAAI Press. 1992.
- [7] C. J. White. **Enterprise Analytics for the Intelligent e-Business**. USA: IBM Press. 2001.
- [8] J. Grabmeier, A. Rudolph. **Techniques of cluster algorithms in data mining version 2.0**. En IBM Deutschland Informations systeme GmbH. GBIS (Global Business Intelligence Solutions). Germany. IBM Press. 1998.
- [9] C. Baragoin, R. Chan, H. Gottschalk, G. Meyer, P. Pereira, J. Verhees. **Enhance your business applications. Simple integration of advanced data mining functions**. En IBM International Technical Support Organization. USA, IBM Press. 2002.
- [10] Ch. Ballard, J. Rollins, J. Ramos, A. Perkins, R. Hale, A. Dorneich, E. Cas Milner, J. Chodagam. **Dynamic warehousing: Data mining made easy**. En IBM International Technical Support Organization. USA, IBM Press. 2007.
- [11] Ch. Ballard, A. Beaton, D. Chiou, J. Chodagam, M. Lowry, A. Perkins, R. Phillips, J. Rollins. **Leveraging DB2 data warehouse edition for business intelligence**. En IBM International Technical Support Organization. USA, IBM Press. 2006.
- [12] J. Gutierrez. **Data Mining, Extracción de Conocimiento en Grandes Bases de Datos**. Ponencia o documento presentado en VII Congreso Argentino de Ciencias de la Computación – V Escuela Internacional de Informática, Calafate, Argentina. 2001.
- [13] IBM Software Group. **Enterprise Data Warehousing with DB2: The 10 Terabyte TPCB Benchmark**. IBM Press. USA. 2003.
- [14] G. Salton. **Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer**. Addison-Wesley. 1989.
- [15] W. Inmon. **Data Warehouse Performance**. USA: John Wiley & Sons. 1992.

- [16] W. Inmon. **Building the Data Warehouse**. USA: John Wiley & Sons. 1996.
- [17] J. Trujillo, M. Palomar, J. Gomez. **Applying Object-Oriented Conceptual Modeling Techniques To The Design of Multidimensional Data-bases and OLAP Applications**. En First International Conference On Web-Age Information Management WAIM 2000. Lecture Notes in Computer Science 1846. 83-94. 2000.
- [18] A. Holzinger, I. Jurisica. **Knowledge Discovery and Data Mining in Biomedical Informatics: The Future Is in Integrative, Interactive Machine Learning Solutions**. Springer-Verlag. LNCS 8401. Knowledge Discovery and Data Mining in Biomedical Informatics. 1-18. 2014.
- [19] K. Srinivas, B. Kavihta Rani, A. Govrdhan. **Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks**. International Journal on Computer Science and Engineering. 2-2. 250-255. 2010.
- [20] H. C. Koh, G. Tan. **Data Mining Applications in Healthcare**. Journal of Healthcare Information Management. 19-2. 64-72. 2005.
- [21] I. M. Mullins, M. S. Siadaty, J. Lyman, K. Scully, C. T. Garrett, W. G. Miller, R. Muller, B. Robson, C. Apte, S. Weiss, I. Rigoutsos, D. Platt, S. Cohen, W. A. Knaus. **Data mining and clinical data repositories: Insights from a 667,000 patient data set**. ELSEVIER. Computers in Biology and Medicine. 36. 1351–1377. 2006.
- [22] N. Bhatla, K. Jyoti. **An Analysis of Heart Disease Prediction using Different Data Mining Techniques**. International Journal of Engineering Research & Technology. 1-8. 1-4. 2012.
- [23] J. Soni, U. Ansari, D. Sharma. **Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction**. International Journal of Computer Applications. 17-8. 43–48. 2011.
- [24] D. Tomar, S. Agarwal. **A survey on Data Mining approaches for Healthcare**. SERSC. International Journal of Bio-Science and Bio-Technology. 5-5. 241-266. 2013.
- [25] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, I. Chouvarda. **Machine Learning and Data Mining Methods in Diabetes Research**. ELSEVIER. Computational and Structural Biotechnology Journal. 15. 104–116. 2017.
- [26] J. F. Easton, C. R. Stephens, M. Angelova. **Risk factors and prediction of very short term versus short / intermediate term post-stroke mortality: A data mining approach**. ELSEVIER. Computers in Biology and Medicine. 54. 199–210. 2014.
- [27] C. Colak, E. Karaman, M. Gokhan TurtaybaI. **Application of knowledge discovery process on the prediction of stroke**. ELSEVIER. Computer Methods and Programs in Biomedicine. 119. 181–185. 2015.
- [28] A. K. Arslan, C. Colak, M. E. SarihanbaInonu. **Different medical data mining approaches based prediction of ischemic stroke**. ELSEVIER. Computer Methods and Programs in Biomedicine. 130. 87–92. 2016.
- [29] V. Paramasivam, T. S. Yee, S. K. Dhillon, A. S. Sidhu. **A methodological review of data mining techniques in predictive medicine: An application in hemodynamic prediction for abdominal aortic aneurysm disease**. ELSEVIER. Biocybernetics and Biomedical Engineering. 34. 139–145. 2014.
- [30] H. Alinejad-Rokny, E. Sadroddiny, V. Scaria. **Machine learning and data mining techniques for medical complex data analysis**. ELSEVIER. Neurocomputing. 276. 1. 2018.
- [31] A. Kalantari, A. Kamsin, S. Shamshirband, A. Gani, H. Alinejad-Rokny, A. T. Chronopoulos. **Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions**. ELSEVIER. Neurocomputing. 276. 2–22. 2018.

- [32] J. Falip, A. Aït-Younes, F. Blanchard, B. Delemer, A. Diallo, M. Herbin. **Visual instance-based recommendation system for medical data mining**. ELSEVIER. Procedia Computer Science. 112. 1747–1754. 2017.
- [33] A. Morais, H. Peixoto, C. Coimbra, A. Abelha and J. Machado. **Predicting the need of Neonatal Resuscitation using Data Mining**. ELSEVIER. Procedia Computer Science. 113. 571–576. 2017.
- [34] R. Peixoto, L. Ribeiro, F. Portela, M. F. Santos and F. Rua. **Predicting Resurgery in Intensive Care - A data Mining Approach**. ELSEVIER. Procedia Computer Science. 113. 577–584. 2017.
- [35] M. K. Obenshain. **Application of Data Mining Techniques to Healthcare Data**. MAT. Statistics for Hospital Epidemiology. 25-8. 690-695. 2004.
- [36] H. Kaur, S. K. Wasan. **Empirical Study on Applications of Data Mining Techniques in Healthcare**. Science Publications. Journal of Computer Science. 2-2. 194-200. 2006.
- [37] D. L. la Red Martínez, S. M. Gerzel, C. E. Aguirre, R. A. Cáceres, J. A. Berecoechea Galarza. **Uso de Minería de Datos Para la Determinación de Perfiles Socioeconómicos y Sanitarios en la UNCAus**. European Scientific Journal (ESJ). Volume 16 – N° 33. pp. 88-104. ISSN N° 1857-7881. University Ss "Cyril and Methodius" Skopje, Macedonia. 2020.
- [38] D. L. la Red Martínez, S. M. Gerzel, C. E. Aguirre, R. A. Cáceres, J. A. Berecoechea Galarza, S. Vallejo. **Perfiles socioeconómicos y sanitarios de barrios de Sáenz Peña, Chaco - Ciencia, tecnología y sociedad unidas por la universidad**. Editorial Académica Española - OmniScriptum S.R.L. Publishing Group. ISBN N° 978-620-388625-2. 2022.

VISUALIZANDO INFORMACIÓN DE EVENTOS CLIMÁTICOS MEDIANTE PRESENTACIONES INTERACTIVAS UTILIZANDO KNIME ANALYTICS

Raúl Oscar Klenzi, Maria Isabel Masanet, Martín Muñoz

Instituto de Informática / Departamento de Informática /Facultad de Ciencias Exactas
Físicas y Naturales / Universidad Nacional de San Juan

Av. Ignacio de la Roza 590 (O), Complejo Universitario "Islas Malvinas", Rivadavia, San Juan, Teléfonos:
4260353, 4260355 Fax 0264-4234980, Sitio Web: <http://www.exactas.unsj.edu.ar>
e-mail: {rauloscarklenzi, mimasanet, martinaz0101}@gmail.com

RESUMEN

La propuesta presenta el estado de avance y logros del proyecto “Evaluación de visualizaciones eficientes en Ciencia de Datos” durante el año 2021 y que como objetivo central pretende analizar y evaluar diferentes alternativas de visualización en lo concerniente al conocimiento extraído desde datos en el marco de la Ciencia de Datos (-Data Science-DS). Se presentan alternativas de visualización interactivas de información y conocimiento extraído de datos provistos por dos estaciones meteorológicas, inherentes a eventos climáticos como detección de heladas y particularmente viento zonda, utilizando para ello el entorno de software licencia GNU, KNIME ANALYTICS (KA). La utilización de gráficos interactivos conforme a la información presentada, facilitan la comprensión e interrelación entre diferentes métricas de valoración de modelos de Aprendizaje de Máquina (Machine Learning – ML-) utilizados en la predicción de los fenómenos.

Palabras clave: Visualización, Ciencia de Datos, Machine Learning, Knime Analytics, Gráficos Interactivos.

CONTEXTO

El proyecto “Evaluación de visualizaciones eficientes en Ciencia de Datos” comienza su tercer año de ejecución en el ámbito del

Departamento de Informática y el Laboratorio de Sistemas Inteligentes para Extracción de Conocimiento en Datos Masivos perteneciente al Instituto de Informática de la Facultad de Ciencias Exactas Físicas y Naturales DI-Idel—FCEFyN vinculando sinérgicamente el accionar de docentes investigadores y alumnos pertenecientes a las carreras de LSI y LCC y cuenta con el aval y subsidio del Consejo de Investigaciones Científicas y Técnicas y de Creación Artística CICITCA-UNSJ. La propuesta cuenta con antecedentes logrados en el tema conforme a sucesivos proyectos aprobados y subsidiados por el ente mencionado en los que el grupo, conformado por docentes responsables de diferentes asignaturas que dan soporte al área del DS, viene trabajando desde el año 2003.

1. INTRODUCCIÓN

La Ciencia de Datos es el campo interdisciplinario tendiente a extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, ya sean estructurados o no estructurados, proponiendo para ello diferentes métricas y formas de análisis y visualización.

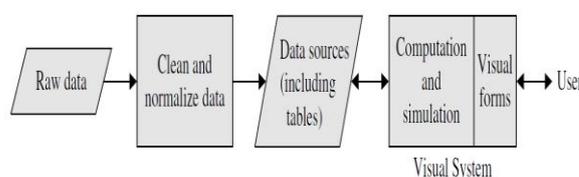


Figura 1: Pipeline de un proceso de DS.

En este caso se propone trabajar y evaluar un determinado modelo de predicción utilizando el módulo de visualización *Binary Classifications Inspector* provisto por la plataforma de DS, Knime Analytics.

La misma, es una conocida plataforma de aprendizaje automático, gratuita, de código abierto que proporciona análisis, integración e informes de datos de extremo a extremo del pipeline. Con la Plataforma KA, los científicos de datos habilitan fácilmente la creación de flujos de trabajos visuales desde una interfaz gráfica, arrastrando y soltando los diferentes nodos que permiten la construcción del workflow de DS. No requiere conocimientos de codificación. Para crear estos flujos de trabajo, un usuario puede elegir entre más de 2000 nodos. KA permite a los desarrolladores realizar varias acciones; desde E/S básica hasta manipulaciones de datos, transformaciones y minería de datos. Knime Analytics consolida todo el proceso en un solo workflow.

El módulo *Binary Classifications Inspector* tiene entre otras, las siguientes propiedades: Permite Comparar una serie de modelos de aprendizaje automático de clasificadores binarios que predicen el mismo objetivo con igual conjunto de datos utilizando métricas de rendimiento y curvas ROC. Así mismo, el usuario puede mejorar la performance de un modelo encontrando el mejor umbral dada una métrica de rendimiento de su elección, tanto como seleccionar de forma interactiva, uno de los valores predichos.

1.1. Consideraciones sobre el problema abordado.

En la provincia de San Juan, particularmente en el Valle del Tulum, la zona agroproductiva que allí se desarrolla se ve amenazada por dos fenómenos meteorológicos adversos como lo son la ocurrencia de las *heladas tardías* en primavera y de viento *Zonda* en el momento de floración y cuaje. Esto afecta en mayor medida

a los cultivos de tipo frutales y en menor medida a cultivos de tipo hortícolas.

El viento, provoca un descenso brusco de la humedad relativa que llega a valores muy bajos (cercanos a cero), y un incremento notable de la temperatura, pudiendo llegar a más de 45° C. Esto produce una marcada deshidratación de los órganos florales, principalmente de los estigmas, evitando la adherencia y germinación del grano de polen y posterior fecundación del óvulo. (Golberg, 2010).

Es de gran interés (en el orden local) realizar el monitoreo y predicción de estos eventos climáticos. En particular la predicción del viento zonda, permite activar mecanismos de defensa, planificar barreras de protección, activar sistemas de riego que contrarresten el descenso de humedad, entre otros.

Los datos a procesar, con el objeto de predecir el evento meteorológico del viento zonda, se presentan en la Tabla 1. Allí se observan datos de entrada asociados a la presión atmosférica, Humedad, temperatura, velocidad del viento, punto de rocío y la presencia o ausencia de viento zonda como variable objetivo.

Row ID	D bar	D hum_ext	D temp_ext	D station	D vel_viento	D pto_rocio	I month	I hour_z...	S isZonda
Row59	419	-3.1	1.53	200	1.6	5	6	1	YES
Row60	400	-2.6	1.18	200	0	0.6	8	3	YES
Row61	-960	5	-1.83	200	0	-1.6	8	16	YES
Row62	-910	-6.3	2.34	200	0	1.3	9	5	YES
Row63	-790	-3.1	1.76	200	3.2	0.5	11	11	YES
Row64	-1,120	4.2	3.25	200	4.8	-1.3	6	3	YES
Row65	-440	3.6	-1.41	200	0	-1.4	7	13	YES
Row100	107	-0.9	0.33	200	6.4	0.1	7	23	NO
Row102	41	2.3	-1.18	200	0	-0.1	8	13	NO
Row103	-21	2.1	-0.79	200	0	1.9	9	12	NO
Row104	85	5	-0.89	200	3.2	6	10	1	NO
Row105	120	-0.9	0.86	200	1.7	-3.9	9	17	NO
Row106	-32	-4.9	1.53	200	1.6	-4.7	6	9	NO
Row107	-30	-1.1	0.17	200	0	-1	5	9	NO
Row108	55	1.1	-0.09	200	-1.6	-1.3	7	3	NO
Row109	67	0.2	-0.47	200	0	1.2	6	21	NO

Tabla1: Visualización de Datos de entrada asociados el evento climático viento zonda en Knime Analytics

Atento a que la variable objetivo es bivaluada se pueden utilizar, entre otros, algoritmos de aprendizaje de máquina como: Random Forest, Decisión Tree, K-Vecinos, etc

En la presente propuesta se presenta la estrategia de los K-Vecinos (con dos valores de K). En este tipo de aprendizaje se entrena al algoritmo con un conjunto de datos que consisten en pares, donde una componente del par serán los datos de entrada (un vector o punto de datos) y la otra los resultados deseados, esta última es denominada atributo de clase. El objetivo del aprendizaje es obtener una función o modelo de clasificación que relacione los datos de entrada con el atributo de clase, y usar así la función para predecir los valores del atributo de clase (Bing, 2007).

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las líneas de investigación consideradas son las atinentes al tratamiento de datos y su procesamiento con el objetivo de predecir variables de interés y con ello tomar las correspondientes decisiones desde formas de visualización de conclusiones ágiles, didácticas y efectivas y que pueden resumirse en las siguientes:

- Ciencia de Datos, principalmente lo relativo a Visualización de Información.
- Deep Learning a través del análisis de Redes Neuronales del tipo LSTM en diferentes entornos de visualización.
- Prototipado de aplicaciones en diferentes herramientas de software libre para arquitecturas secuenciales, paralelas y distribuidas particularmente KNIME Analytics cuyo enorme potencial didáctico permite visualizar mediante workflows las diferentes etapas del pipeline del Data Science.
- Desarrollos finales en lenguajes de programación de código abierto tales como Python, JavaScript y R.
- Paralelización de algoritmos que desde la habilitación de parámetros de las herramientas y lenguajes de software antes mencionados, permitan la

utilización de placas gráficas del tipo GPU (Graphics Processing Unit).

3. RESULTADOS OBTENIDOS Y ESPERADOS

Se encuentran en estado de próxima defensa, dos trabajos finales de grado que tienen que ver con la monitorización, relevamiento y visualización de variables climáticas propias de la zona centro oeste de la Argentina en diferentes plataformas de hardware. En este contexto, en lo atinente al fenómeno climático del viento zonda y como una alternativa de utilización del módulo *Binary Classifications Inspector*, la Figura 1 presenta el pipeline que desde datos relevados por dos estaciones meteorológicas marcan la presencia o ausencia de viento zonda. Subsiguientes módulos separaan los datos en un conjunto de entrenamiento y y en otro de testeo con el objetivo de evaluar un clasificador basado en el algoritmo de los K-vecinos, para finalmente contrastar para valores de 3 y 5 vecinos respectivamente los modelos considerados

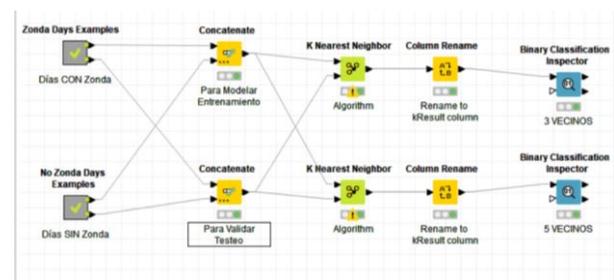


Figura 1: Pipeline del clasificador mediante K-Vecinos para la predicción de viento zonda.

Las Figura 2 y 3, evidencian la riqueza en cuanto a didáctica, versatilidad, visualización e interacción de métricas asociadas al modelo de ML utilizado, al observar y analizar la información provista por el módulo *Binary Classifications Inspector*.

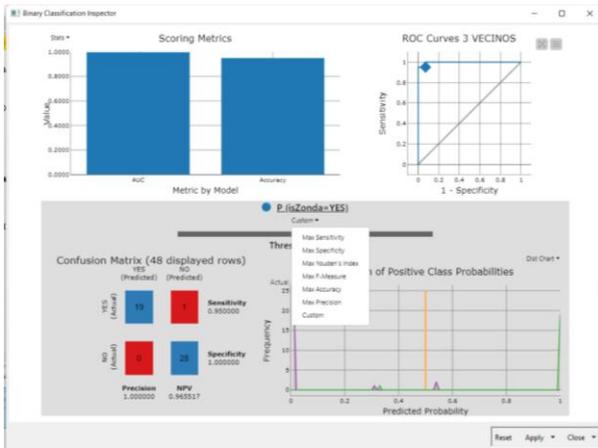


Figura 2: Visualización de Información para la validación de resultados correspondiente al modelo de 3-Vecinos.

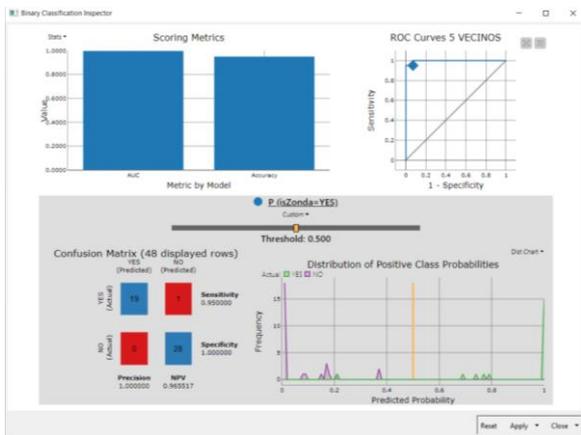


Figura 3: Visualización de Información para la validación de resultados correspondiente al modelo de 5-Vecinos.

El módulo *Binary Classifications Inspector* permite, entre otras cosas, comparar el Área Bajo la Curva (Area Under Curve –AUC–) del modelo para seleccionar su mejor versión a través del gráfico de barras de estadísticas y/o curvas ROC asociadas al mismo.;

- Cambiar el umbral desde su valor inicial, manualmente o a través de su control deslizante, o maximizando automáticamente una de las funciones disponibles a través del menú desplegable de la pestaña del Modelo y reconocer su interacción contras métricas.

- Analizar la Matriz de confusión en el panel inferior izquierdo, para evaluar la gravedad de una clasificación errónea.

Desde estos análisis se observa que el área bajo la curva en la vista superior derecha, los niveles de exactitud parte superior izquierda de la imagen y la propia matriz de confusión reflejan una instancia de modelación más que promisoria para los datos de entrada y para ambas instancias de K-Vecinos consideradas con un umbral de 0,5.

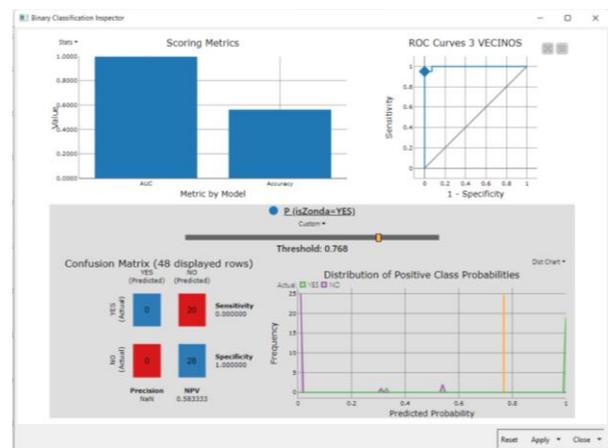


Figura 4: Visualización de Información correspondiente al modelo de 3-Vecinos tras modificación de umbral.

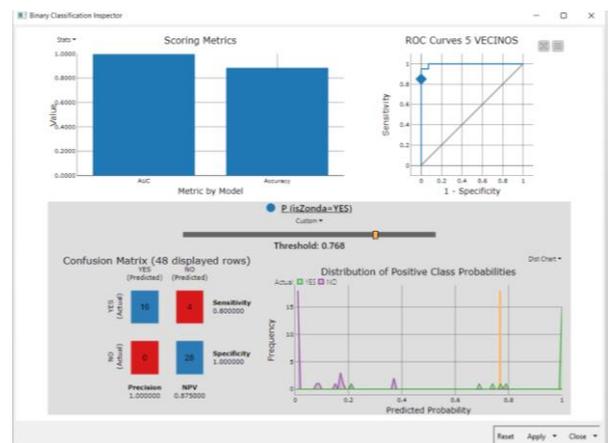


Figura 5: Visualización de Información correspondiente al modelo de 5-Vecinos tras modificación de umbral.

Se puede observar sin embargo, en las Figuras 4 y 5 y dada la posibilidad de interactuar en el entorno gráfico, que para un mismo nivel

variación de umbral el modelo 5-vecinos, refleja mayor robustez que el de 3-vecinos.

Se ha presentado una comparación entre dos modelos de K-vecinos con diferentes valores de parámetros internos, lo mismo se puede realizar para diferentes algoritmos y es lo que se intentará plasmar desde la prototipación de aplicaciones en el entorno Knime durante el presente 2022 y trasladar la mejor versión de los modelos a aplicaciones desarrolladas en lenguaje Python o Java.

4. FORMACIÓN DE RECURSOS HUMANOS

El proyecto de investigación está integrado por docentes investigadores responsables de cátedras de las carreras del DI-FCEFyN y que son inherentes a las diferentes columnas que dan soporte y conforman el área de conocimiento de la Ciencia de Datos. Estos Docentes investigadores fomentan constantemente la participación de alumnos que integran el espacio e incitan a aquellos que cursan esas cátedras a abrazar la temática e integrarse al proyecto. Durante el pasado 2021, se han alcanzado los siguientes logros:

- Culminación y defensa de la propuesta “Herramienta tecnológica de apoyo al aprendizaje: Problema del Viajante de Comercio, caso asimétrico” (Autor: Cocinero, Pablo. Defendido exitosamente en Marzo del 2021 y presentado en TE&ET del mismo año).
- El trabajo llevado adelante en el marco de la Tesis de Maestría en Informática "Análisis de Fenómenos en Estaciones Agrometeorológicas mediante Ciencia de Datos" a cargo de una integrante del equipo de trabajo permite en este momento, encontrarnos ante las últimas revisiones de dos informes de trabajo final de Licenciatura en Ciencias de la Computación del DI-FCEFyN, con próxima defensa de los mismos y en los cuales se ha logrado una fuerte interacción y sinergia con

proyectos de investigación aplicada llevada adelante en dependencias del UNAUT-FI-UNSJ y con el INTA-SJ organización proveedora de los datos procesados. Así mismo permitió exponer en el XIX RPIC 2021 el trabajo investigativo “Técnicas De Balanceo De Datos Para Predecir La Ocurrencia Del Fenómeno Meteorológico De La Helada”

- La elevación del informe final correspondiente a la Beca CIN “Abordaje de la Analítica Visual desde un lenguaje de programación- Python como caso de estudio” otorgada en Agosto de 2020 y ejecutada hasta Julio de 2021, permite actualmente concretar, para el mismo alumno que ejecutó la mencionada beca, una propuesta de trabajo final que continuará y profundizará el conocimiento sobre la temática de aplicación de los módulos de python asociados al área de visualización en Ciencia de Datos orientada al análisis de falsos positivos en la determinación de particulares eventos climáticos como la helada a efectos de minimizar costos que tienen que ver con la activación de acciones de mitigación ante la ausencia del fenómeno real.

5. BIBLIOGRAFÍA

- Glenn J. Myatt, Wayne P. Johnson-Making Sense of Data III_ A Practical Guide to Designing Interactive Data Visualizations -Wiley (2011)
- Masanet M., Capraro F., Klenzi, R., Muñoz, M.. Procesamiento de Datos Meteorológicos Para Determinar La Ocurrencia de Heladas en Agricultura. (XXIII WICC 2021, Chilecito, La Rioja).
- Tshepo Chris Nokeri (2021). Data Science Revealed. With Feature Engineering, Data Visualization, Pipeline Development, and Hyperparameter Tuning. Apress

Propuesta de técnicas de validación para la calidad de datos abiertos e identificación de patrones para predicciones con Machine Learning

Roxana Martínez, Christian Parkinson, Martín Caruso, Diego López,
Rocío Vargas, Nayiby Rojas

CAETI - Centro de Altos Estudios en Tecnología Informática
Universidad Abierta Interamericana (UAI)
Montes de Oca 745, Ciudad Autónoma de Buenos Aires, Argentina

{roxana.martinez, christian.parkinson, rocio.vargas}@uai.edu.ar

{martin.caruso, diego.lopez, nayibi.rojas}@alumnos.uai.edu.ar

RESUMEN

La política de los datos abiertos busca promover la innovación y transformar la actividad gubernamental para brindar mejores servicios y generar mayores niveles de transparencia en la sociedad. Por lo que, mantener la calidad en las fuentes de datos disponibilizadas es fundamental para su tratamiento y obtener así, un conocimiento de éstas. Actualmente, son pocos los trabajos realizados en aspectos de validaciones, análisis de contenidos internos de estos datasets, herramientas de datos, identificación de patrones en su estructura y demás. En base a esto, esta línea de investigación se enfoca en el análisis, diseño y desarrollo de herramientas de software que utilicen técnicas y propuestas para la validación de la calidad de los datos públicos abiertos en el contexto de Gobierno Abierto. Además de detectar el “estado de salud” de estos datos (grado de integridad, redundancia y otros), se trabaja en el estudio de patrones con el fin de incorporar modelos de predicción para aportar un valor agregado a la información otorgada por los gobiernos desde aspectos de Machine Learning, lo que ofrecerá un mejor análisis para la toma de decisiones gubernamentales.

Palabras clave: Datos Abiertos, Gobierno Abierto, Métricas de Calidad de Datos, Machine Learning, Predicciones a partir de Patrones.

CONTEXTO

El presente trabajo es parte del proyecto denominado “Investigación y desarrollo de software para la validación de la calidad de datos abiertos e identificación de patrones para predicciones”, que tiene inicio en el mes de marzo 2022. Este proyecto pertenece a la línea de investigación de Ingeniería de Software del Centro de Altos Estudios en Tecnología Informática (CAETI) de la Facultad de Tecnología Informática de la Universidad Abierta Interamericana (UAI), el cual contribuye al desarrollo de las Tecnologías de la Información y las Comunicaciones (TIC's) en Argentina y en el mundo, llevando adelante la investigación básica y aplicada en diversas áreas. El proyecto es financiado y evaluado por la Secretaría de Investigación de la Universidad, tiene una duración de 2 años, y cuenta con la participación de docentes y estudiantes de grado y posgrado en diversas carreras de la Facultad de Tecnología Informática.

1. INTRODUCCIÓN

El Gobierno Abierto se enfoca en la “construcción de estados transparentes, participativos, que rindan cuentas de manera adecuada e innovadores, poniendo al ciudadano en el centro de la toma de las decisiones públicas, como una forma de fortalecer el Estado democrático” [1]. Este contexto implica la utilización de diversas tecnologías abiertas con el fin de fomentar la

Innovación Pública. Por otra parte, se sostiene que este término “es una doctrina política que surge a partir de la adopción de la filosofía del movimiento del software libre a los principios de la democracia. Este paradigma tiene como objetivo que la ciudadanía colabore en la creación y mejora de servicios públicos y en el robustecimiento de la transparencia y la rendición de cuentas” [2]. Por lo que el Estado Nacional debe incorporar las técnicas para gestionar dicha cantidad de datos con un diseño tecnológico y por sobre todo enfocado en el ciudadano promedio para fomentar la inclusión social. “Al ampliar el acceso a la información pública se fortalece la rendición de cuentas y se enriquece el debate público, a la vez que se crean nuevas oportunidades para generar valor agregado” [3]. El enfoque de formato abierto es el formato de archivo no propietario, cuya especificación debe estar documentada públicamente, es de libre conocimiento e implementación y libre de patentes o de cualquier otra restricción legal o económica para su uso. Para ello, es necesaria una licencia abierta, que es un acuerdo de provisión de datos para que cualquier persona los utilice, los reutilice y los distribuya, estando sujeto a las condiciones de dicha licencia.

“Los datos que se pueden reutilizar y redistribuir sin ninguna restricción se denominan datos abiertos” [4]. Estos datos deben estar en formatos digitales, con un modelo estándar abierto. Por otra parte, es importante destacar que, “no toda la información pública disponible o publicada en la web es información abierta válida para su reutilización. No sólo se trata de publicar los datos, sino que hay que garantizar el acceso a ellos, razón por la que debe recurrirse a formatos digitales, estandarizados y abiertos, siguiendo una estructura clara que permita su comprensión y reutilización” [5]. La gestión de la información en formatos abiertos, “datos abiertos”, consiste en el acceso y uso de la información pública por parte de terceros para entregar nuevos servicios a los ciudadanos, esto permite acceder a una gran cantidad de datos procedentes de diferentes organizaciones del ámbito de la

administración pública [6].

Por lo anteriormente explicado, es fundamental que personal dedicado trabaje en el tratamiento de las fuentes de datos abiertos, para que los organismos estatales y ciudadanos, tengan un mejor conocimiento sobre un determinado tema público, como ser: economía, transporte, etc.

Magallón afirma que “la cultura de datos abiertos trata de obtener un valor añadido de la información. A diferencia de lo ocurrido hasta ahora esta información no genera sólo su valor por estar reservada a unos pocos, sino que lo hace por su disponibilidad para ser interpretada y traducida por cualquier actor interesado en trabajar con ella” [7], es decir, un dato abierto es aquel que puede ser accedido, y conlleva un formato que permite la interoperabilidad con otros softwares. Diversos organismos estatales que ofrecen una gran cantidad de fuentes de datos de varios temas gubernamentales con criterios preestablecidos en sus portales, que brindan datasets que son utilizados como insumo fundamental de información y servicios.

Existen algunos trabajos enfocados a las mediciones en aspectos de calidad [8] [9] [10], por ejemplo, el Barómetro de Datos Abiertos, ODB [11] de la World Wide Web Foundation, es una medida global de cómo los gobiernos publican y utilizan los datos abiertos para la rendición de cuentas, la innovación y el impacto social. Otras mediciones fueron desarrolladas por The Global Open Data Index [12] que es el punto de referencia mundial anual para la publicación de datos gubernamentales abiertos, gestionado por la Open Knowledge Foundation [13], funciona como una encuesta de crowdsourcing, que mide la apertura de los datos gubernamentales a través de la metodología GODI [14]. Otro proyecto es el Open Government Data de la Organisation for Economic Co-operation and Development, OECD [15], su objetivo es avanzar en la evaluación de impacto de Open Government Data (OGD), para este caso, su índice evalúa los esfuerzos de los gobiernos para implementar datos abiertos en las tres áreas críticas: apertura, utilidad y reutilización de

los datos gubernamentales; otras medidas son las puntuaciones otorgadas a la calidad de los datos abiertos, este es el caso del esquema modelo de cinco estrellas de Berners-Lee [16], o informes sobre el estado de los datos elaborados por fundaciones comprometidas con el impacto de estos datos públicos [17].

Gracias al gran impulso de la tecnología en el área de datos, el concepto de Machine Learning cada vez se encuentra más en auge, “el aprendizaje automático se refiere al proceso por el cual los ordenadores desarrollan el reconocimiento de patrones o la capacidad de aprender continuamente y hacer predicciones basadas en datos tras lo cual realizan ajustes sin haber sido programados específicamente para ello. Como forma de inteligencia artificial, el aprendizaje automático automatiza el proceso de creación de modelos analíticos y permite que las máquinas se adapten a nuevas situaciones de manera independiente” [18]. Este paradigma, tiene implicaciones para el descubrimiento científico, ya que la complejidad de los patrones que las máquinas son capaces de identificar no es fácilmente lograda por los procesos cognitivos humanos [19].

Machine Learning apunta a la utilización de algoritmos que analizan datos y, a partir de éstos, logran determinar el comportamiento del software. Para lograr un correcto análisis, son necesarios algoritmos que sean alimentados por datos que sostengan estos sistemas automatizados, es decir, una mayor disponibilidad y calidad de los datos abiertos servirá para alimentar esos algoritmos y a la vez poder también mejorarlos y auditar su correcto funcionamiento [20]. “Uno de los puntos importantes para tener en cuenta en el proceso de Machine Learning es el preprocesado y preparación de las variables que componen los conjuntos de entrenamiento y test de algoritmos, ya que será una condicionante esencial” [21] para un buen análisis.

En lo relativo al dato, es necesario contar con técnicas de validación de calidad, debido a que hoy por hoy, son pocos los trabajos realizados en aspectos de enfoque [22], por lo que, contar con las diversas técnicas de

Machine Learning y que éstas puedan ser utilizadas en datos públicos, podría generar un gran beneficio a la sociedad, y, por otro lado, mantener la calidad y la apertura de los datos públicos, ayudará a los gobiernos y a los diferentes actores de la sociedad civil a tomar mejores decisiones, ya que tienen una visión e información de la realidad más precisa.

Uno de los problemas actuales es que, en los portales de datos abiertos, la disponibilidad de los datos no necesariamente coincide con que tengan calidad, lamentablemente, hoy sigue siendo una dificultad y es un gran desafío para las políticas públicas. El análisis de muchos de los conjuntos de datos públicos representa un problema crucial, ya que está disperso, no estandarizado y en muchos casos desactualizado.

2. LÍNEAS DE INVESTIGACIÓN DESARROLLO

Este proyecto pertenece a la línea de investigación de Ingeniería de Software del Centro de Altos Estudios en Tecnología Informática (CAETI). Los ejes principales del tema que se está investigando son:

- Analizar las falencias actuales en cuestiones de calidad de datos abiertos públicos.
- Identificar las mejores técnicas de validación de calidad de datos abiertos disponibilizados.
- Diseñar y desarrollar algoritmos para identificar patrones en datos públicos.
- Analizar modelos predictivos orientados a los datos abiertos.
- Efectuar predicciones sobre nuevos datos encontrados con técnicas de Machine Learning en datos públicos.

3. RESULTADOS OBTENIDOS/ESPERADOS

Enfoque del proyecto:

Los datasets abiertos gubernamentales en contexto de gobierno abierto, tienen falencias en aspectos de calidad de datos, éstos pueden ser detectados y analizados a través de

herramientas de validación y medición para mejorar su interoperabilidad a nivel software, como así también, pueden utilizar algoritmos para identificar patrones y realizar predicciones para lograr beneficios orientados a la comunidad social.

Objetivos principales:

Realizar el análisis, diseño y desarrollo de herramientas de software para la gestión y validación de la calidad de los datos públicos en el contexto de Gobierno Abierto. Detectando el “estado de salud” de las diversas fuentes de datos provenientes de casos de aplicación gubernamentales con los prototipos desarrollados, incorporando algoritmos para identificar patrones que logren predicciones sobre nuevos datos.

Objetivos específicos esperados:

Estudiar el alcance de los conceptos implicados en este contexto de estudio; Analizar los trabajos relacionados en la temática de Gobierno Abierto relativos al tratamiento de datos abiertos y públicos; Analizar las técnicas de Machine Learning para datos abiertos; Relevar los últimos trabajos relacionados en cuanto a los aspectos de calidad de datos abiertos y públicos; Relevar trabajos enfocados a Técnicas de Machine Learning con datos abiertos; Analizar los criterios de los portales de datos abiertos para efectuar la publicación de éstos; Relevar los distintos tipos y formatos disponibles de datos abiertos que existen en los dataset actuales de los sitios más relevantes y gubernamentales de Argentina; Analizar las falencias de los datasets disponibles en los sitios de Gobierno; Analizar herramientas de manejo con Machine Learning que permitan análisis de patrones con enfoque predictivo; Establecer criterios estándar de calidad de datos; Desarrollar una herramienta de validación de datasets gubernamentales basada en métricas de calidad de datos; Generar una guía de buenas prácticas para las técnicas de Machine Learning en la utilización de datos abiertos; Definir y recolectar la muestra de datasets para ser testeados por la herramienta

propuesta como validadora de calidad; Análisis de casos predictivos con casos reales; Análisis de casos de aplicación con los prototipos desarrollados.

Metodología y Técnicas:

El proceso metodológico de investigación empleado para el presente proyecto de investigación se define con un proceso sistemático cualitativo, que implementa una forma evolutiva incremental en cada una de las etapas involucradas, siendo estas: identificación del problema; revisión teórica; recolección de datos, clasificación y análisis de datos; estudio de escenarios de identificación de patrones sobre éstos con enfoques predictivos sobre nuevos datos; desarrollo de prototipos (análisis cuantitativo para cada una de las métricas de calidad de datos); visualización de la predicción de patrones; validación de la solución; identificación de las limitaciones del trabajo.

4. FORMACIÓN DE RECURSOS HUMANOS

Este proyecto se compone por 3 (tres) docentes con estudios de posgrado: uno de ellos Magíster en Tecnología Informática y a la espera de respuesta por parte del jurado que se encuentra en revisión de su tesis doctoral en Ciencias Informáticas en la Universidad Nacional de La Plata (UNLP), y dos docentes que se encuentran realizando su tesis de maestría en Tecnología Informática en la Universidad Abierta Interamericana (UAI). El equipo también cuenta con la participación de estudiantes de grado y de posgrado de la UAI.

En relación directa con la línea de I/D presentada para el proyecto, los miembros del equipo se encuentran en realización de: 1 tesis doctoral (docente), 2 tesis de maestría (un docente-estudiante y un estudiante) y 2 tesinas de grado (estudiantes) en la UAI.

5. BIBLIOGRAFÍA

[1] Arroyo Chacón, J. (2017). La Innovación Abierta Como Pilar Del Gobierno Abierto

- (Open Innovation as a Pillar of Open Government). *Revista Enfoques*, 15(27), 13-41.
- [2] Sosteniblepedia.org. “*Gobierno Abierto*”. Disponible en: https://www.sosteniblepedia.org/index.php?title=Gobierno_abierto
- [3] Buenos Aires provincia (2017). “Kit de Apertura Municipal”. Disponible en: <http://escueladefiscales.com/Kit%20de%20Apertura%20Municipal%202017%20-%20provincia%20de%20buenos%20aires.pdf>
- [4] Oviedo, E., Mazón, J. N., & Zubcoff, J. J. (2013). Hacia un modelo de calidad de datos para portales de datos abiertos. In XXXIX Latin American Computing Conference (CLEI), Nanguata (pp. 1-8).
- [5] Garriga-Portolà, M. (2011). ¿Datos abiertos? Sí, pero de forma sostenible. *Profesional de la Información*, 20(3), 298-303.
- [6] Naser, A., & Ramírez Alujas, Á. (2017). Plan de gobierno abierto: una hoja de ruta para los gobiernos de la región.
- [7] Magallón Rosa, R. (2017). Datos abiertos y acceso a la información pública en la reconstrucción de la historia digital.
- [8] ISO 25012 (2008). “*Ingeniería de software - Requisitos de calidad y evaluación de productos de software (SQuaRE) - Modelo de calidad de datos*”. Disponible en: <https://www.iso.org/obp/ui/es/#iso:std:iso-iec:25012:ed-1:v1:en>
- [9] Martínez, R. et al. (2021). Metrics proposal to measure the quality of governmental datasets. *IEEE Latin America Transactions*, Vol. 100. ISSN 1548-0992.
- [10] de España, G. (2017). Manual práctico para mejorar la calidad de los datos abiertos. Madrid. Disponible en: https://datos.gob.es/sites/default/files/doc/file/manual_practico_para_mejorar_la_calidad_de_los_datos_abiertos_1.pdf
- [11] Open Data Barometer. “*The Open Data Barometer*”. Disponible es: https://opendatabarometer.org/?_year=2017&indicator=ODB
- [12] Open Knowledge Foundation (2020). “*Global Open Data Index - Argentina*”. Disponible en: <https://index.okfn.org/place/ar/>
- [13] Open Knowledge Foundation (2020). “*A fair, free and open future*”. Disponible en: <https://okfn.org/>
- [14] Open Knowledge Foundation Network (2017). “*Methodology - Global Open Data Index*”. Disponible en: <https://index.okfn.org/methodology/>
- [15] OECD Better policies for better lives. “*Open Government Data*”. Disponible en: <http://www.oecd.org/internet/digital-government/open-government-data.htm>
- [16] Open Data (2012). “*5 Open Data*”. Disponible en: <https://5stardata.info/en/>
- [17] ODI Open Data Institute. “*The 2019 Data Skills Framework*”. Disponible en: <https://theodi.org/article/open-data-skills-framework/>
- [18] Herrera Carrasco, J. (2020). Evaluación de Modelos de Transporte mediante datos abiertos y técnicas de Aprendizaje Automático.
- [19] Boulton, G., Hodson, S., Babini, D., Li, J., Marwala, T., Musoke, M. G., ... & Wyatt, S. (2017). Datos abiertos en un mundo de grandes datos: Un acuerdo internacional ICSU-IAP-ISSC-TWAS. *Revista iberoamericana de ciencia tecnología y sociedad*, 12(34), 267-272.
- [20] Datos.gob.es (2017). “*El futuro de los datos abiertos y sus múltiples caras*”. Disponible en: <https://datos.gob.es/es/noticia/el-futuro-de-los-datos-abiertos-y-sus-multiples-caras>
- [21] Gómez, C. E. J., & Roma, J. C. (2018). Análisis predictivo de datos abiertos sobre el uso turístico del servicio de alquiler compartido de bicicletas de Nueva York. Universidad Oberta de Catalunya, Master Universitario en Ciencia de Datos.
- [22] Kumar, V. D., & Alencar, P. (2016, December). Software engineering for big data projects: Domains, methodologies and gaps. In 2016 IEEE International Conference on Big Data (Big Data) (pp. 2886-2895). IEEE.

APLICACION DE INTELIGENCIA Y ANALITICA DE NEGOCIOS EN DIFERENTES CONTEXTOS

Silvina I. Migani¹, María Inés Lund¹, Alejandra Orellana Vassallo¹, Cristina Vera¹, Alejandro Riveros^{1*2}, Martín Manassero^{1*}, María Laura Molina^{1*}, Diego Guevara^{1*}, Juan Capdevila^{1*}, Leandro Drazic^{1*}, Adriana María Gómez², Sonia Eleonora Pinto², Diego Checarelli², Mariano Alaniz^{1*}

¹Departamento de Informática, FCEFN, Universidad Nacional de San Juan

^{1*}Alumno Departamento de Informática, FCEFN, Universidad Nacional de San Juan

²Gobierno de la Provincia de San Juan

silvina.migani@gmail.com, {mlund, oalitaorellana}@unsj-cuim.edu.ar, {civerados, alejandroriveros09, manasseromm, lauramolinaf86, guevaradiego08, juan.capdevila27, leandrodrazic, a.gomeztapella, pinto.eleo, diegochecarelli, mariano.alaniz12}@gmail.com

RESUMEN

La Inteligencia de Negocios (IN) es el área de conocimiento que comprende un variado conjunto de metodologías, procesos y tecnologías que generan información y conocimiento para sustentar la toma de decisiones de cualquier organización. Dentro de ella, se encuentra la Analítica de Negocios (AN), y refiere concretamente a la utilización de técnicas analíticas que permiten alcanzar el propósito de la IN. Su aplicación se ha convertido en un hecho vital, más allá de la dimensión, sector, modalidad o contexto del organismo considerado.

Alineados a los objetivos del proyecto, se han investigado, desarrollado y/o aplicado métodos y herramientas de software en diferentes entornos y dimensiones. Específicamente se ha trabajado en el ámbito educativo, gubernamental, empresarial y de inversiones, entre otros; dando lugar a publicaciones que se exponen en la sección de resultados.

Palabras claves: Inteligencia de Negocios, Analítica de Negocios, Toma de Decisiones.

CONTEXTO

Este trabajo se enmarca en el proyecto “E1131-INTELIGENCIA DE NEGOCIOS: TECNOLOGÍAS, APLICACIONES Y TENDENCIAS”, aprobado por la UNSJ - Res.N°0591-20-R, a desarrollarse durante el período 2020-2022.

El equipo de trabajo del proyecto es variado. Está constituido por personas que forman parte de la UNSJ (docentes y alumnos de grado y postgrado) y otras que se desempeñan en el gobierno provincial y están fuera del ámbito universitario. Los investigadores docentes se desempeñan también en asignaturas (pertenecientes a las dos licenciaturas del Departamento de Informática) que corresponden a la misma área de conocimiento del proyecto, aspecto que beneficia ampliamente ambos espacios. Por otro lado, la participación de personas externas a la UNSJ, permite conocer y abordar investigaciones hacia problemas concretos requeridos por la sociedad actual.

1. INTRODUCCIÓN

Los procesos de negocio de una organización son todas las actividades interdepartamentales y correlacionadas que juntas logran el funcionamiento lógico de una empresa. Los sistemas de información han contribuido a los negocios innovando las formas de desarrollar sus procesos, adoptando nuevas tecnologías y contribuyendo a una economía más globalizada (Rocha, L. M., citado por [1]).

Hatch, citado en [2], manifestó que: *“La Inteligencia de Negocios es la combinación de prácticas, capacidades y tecnologías usadas por las compañías para recopilar e integrar la información, aplicar reglas del negocio y asegurar la visibilidad de la información en función de una mejor comprensión del mismo y, en última instancia, para mejorar el desempeño”*. Dentro de ella se enmarca la AN, que refiere concretamente a la aplicación de técnicas analíticas que provean esa información mencionada [3].

La IN no sólo permite entender el estado actual del negocio sino también hacer proyecciones a futuro. Brinda apoyo sistemático y estructurado valioso para la toma de decisiones en todos los niveles (operativo, táctico y estratégico) de la pirámide organizacional. Los tipos de decisiones y las características de la información que necesitan son diferentes, desde decisiones estructuradas (nivel operativo), donde la información es rutinaria, predefinida, programada, detallada, frecuente, histórica, interna, y enfocada; hasta decisiones no estructuradas (nivel estratégico), donde la información que requieren es específica, por excepción, no programada, resumida, poco frecuente, prospectiva, externa y de amplio alcance [4], [5].

El mundo actual, donde la evolución tecnológica es vertiginosa y los datos producidos se caracterizan por ser voluminosos, variados y dispersos, las organizaciones tienen que intensificar sus

esfuerzos y generar estrategias para poder incorporar soluciones de IN que les permita optimizar la utilización de sus recursos, controlar el cumplimiento de sus objetivos y tomar buenas decisiones. (Polo, A, A. M., citado por [1]).

Hace décadas que su aplicación en grandes empresas u organizaciones es esencial, les permite beneficiarse extrayendo información de los datos de sus clientes, de sus proveedores y hasta de sus competidores [6], [7]. Por el contrario, las MiPyMEs generalmente siguen tomando sus decisiones en base a la intuición y a la experiencia [8]; pero empiezan a percibir que para lograr permanecer y crecer en un entorno tan competitivo y variable como el actual, necesitan urgentemente basar sus decisiones con información y/o conocimiento preciso y oportuno [9]. Argentina no está exenta de esta realidad. Es así como no sólo las micro y pequeñas empresas han comenzado a incorporar este tipo de sistemas, sino también el gobierno nacional y provincial a través de planes de modernización [10], [11].

Los datawarehouse (DW) constituyen un componente esencial de la IN. En ellos se coleccionan, transforman, estructuran y almacenan los datos organizacionales para posteriormente ser utilizados en tareas de análisis. Algunos autores referentes expresan que un DW, o almacén de datos en español (AD), no es una mera copia de los datos transaccionales en una plataforma diferente; sino que tiene necesidades, objetivos, clientes y ritmos profundamente diferentes a los sistemas operacionales. Otra característica distintiva es que preservan el contexto histórico de los datos para poder analizar el rendimiento de una organización a lo largo del tiempo [12], [13]. Además, están optimizados para consultas de alto rendimiento, ya que a menudo requieren que se explore un gran volumen de datos. La implementación puede realizarse en diferentes tipos de gestores de

bases de datos. La alternativa relacional es la más utilizada, donde los esquemas típicos son estrella y copo de nieve [12], [13]. Sin embargo, provoca inconvenientes en relación al rendimiento en entornos de BIG DATA, donde la cantidad y variedad de datos es excesiva [14], [15]. En este contexto, el uso de sistemas de gestión de bases de datos NoSQL constituye una buena alternativa de solución. Concretamente, los gestores NoSQL basados en grafos prometen ser altamente beneficiosos [16].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El proyecto marco de este trabajo busca estudiar, analizar, comparar y experimentar técnicas, métodos, tipos de análisis y herramientas de software de IN propicios en variados contextos o escenarios organizacionales. Se mencionan a continuación algunas de las líneas de investigación seguidas:

- Investigación y desarrollo de IN y AN en la educación, como innovación educativa.
- Investigación y desarrollo de IN y AN en la pequeña empresa.
- Investigación y desarrollo de IN y AN aplicando un abordaje no relacional.
- Investigación y desarrollo de IN y AN en el ámbito gubernamental.
- Investigación y desarrollo de IN y AN en el mundo financiero de inversión en criptomonedas.
- Investigación y desarrollo de un modelo que sea una propuesta ágil para generar datawarehouses incrementales.

3. RESULTADOS

En el 2021, se incorporaron dos nuevos alumnos al proyecto, en calidad de adscriptos y tesistas de grado. Debido a la pandemia COVID-19 las reuniones de proyecto fueron virtuales. Dados los diferentes contextos de aplicación y necesidades de cada integrante o grupos de trabajo, se decidió pautar reuniones

parciales o personales. Resultó una buena estrategia ya que nos permitió trabajar mejor y lograr avances de cada subgrupo en forma independiente. De todos modos, se efectuaron algunas reuniones globales que fueron grabadas y compartidas con todo el equipo, posibilitando la visualización asincrónica de aquellos que no pudieron participar.

Cabe mencionar que hacer reuniones virtuales nos permitió experimentar una nueva modalidad, que resultó ser una forma por demás ágil y productiva, permitiendo avanzar en la investigación aplicada.

A continuación se enumeran algunos resultados logrados en el proyecto, durante el año 2021:

En el XXIII Workshop de Investigadores en Ciencias de la Computación – XXIII WICC 2021 (UNChilecito, 15 y 16 de Abril). ISBN: 978-987-24611-3-3. Se presentaron los trabajos de dos proyectos:

- “Inteligencia y Analítica de Negocios para la toma de decisiones en diferentes contextos”. Autores: M.I. Lund, S.I. Migani, C.Vera, A.Orellana, A.María Gómez, S.E.Pinto, A.Riveros, D. Checarella, M.Alaniz, L.Drazic, D.Guevara, M.L.Molina.
- “Análisis de calidad de vida de la población sanjuanina aplicando ciencia de datos”. Autores: M.Herrera, S.Ruiz, S.P.Gonzalez, M.G.Romagnano, M.I. Lund, L.Ganga, M.V.Scheffer, M.Sosa.

En las Jornadas Exactas Investiga, desarrolladas entre el 27 y 28 de mayo de 2021, en la FCEF - UNSJ, se expuso el proyecto:

- “Inteligencia de Negocios: tecnologías, aplicaciones y tendencias – E1131” Autores: M.I.Lund y S.I.Migani.

En el International Virtual Workshop on Business Analytics EUREKA 2021. Ciudad de Juárez, Chihuahua, México, 2 al 4 de Junio. Se presentó y expuso, pero no se publicó, el artículo:

- “Applying Business Intelligence and Analytics to Micro and Small Businesses: An Iterative Model for Datawarehouse Generation”. Autores: S.Migani, M.I.Lund.

En el V Congreso Internacional de Ciencias de la Computación y Sistemas de Información 2021 – CICC SI 2021. (18 y 19 de Noviembre), se presentaron los siguientes artículos (proceedings en elaboración).

- “Inteligencia de Negocios en MiPyMEs: Una propuesta ágil para generar datawarehouses incrementales”. Autores: S.I. Migani, M.I.Lund, A.Orellana.
- “Analítica de interacciones en Github implementada en Neo4j”. Autores: C.Vera, S.Migani. S.Zapata.

En el V Congreso Internacional de Ciencias de la Computación y Sistemas de Información 2021 – CICC SI 2021. (18 y 19 de Noviembre), se presentaron los siguientes artículos cortos - Posters (proceedings en elaboración), con los resultados de los alumnos adscriptos y tesisas:

- “Sistema de Inteligencia de Negocios en la Dirección Provincial de Aeronáutica, para el apoyo al proceso de toma de decisiones”. Autores: A.Riveros, M.I.Lund, S.I. Migani.
- “Prototipo desarrollado en Knime para estimar el valor de venta de una criptomoneda”. Autores: D.Guevara, J.Capdevila, S.I.Migani, M.I.Lund.
- “Inteligencia de Negocios para analizar los aspirantes de las carreras de la FCFN-UNSJ”. Autores: M.L.Molina, M.I.Lund, S.I. Migani.
- “Sistema Analítico y Gestor de Cartera de Inversiones en Criptomonedas”. Autores: M.Manassero, M.I.Lund, A.Orellana.

4. FORMACIÓN DE RECURSOS HUMANOS

En términos generales, el proyecto está focalizado a la profundización y consolidación del conocimiento en el área de la analítica e inteligencia empresarial, en diferentes contextos de aplicación, por parte de cada uno de sus integrantes docentes, profesionales y

alumnos. Sin embargo, dentro del grupo de investigación se distinguen los siguientes:

En relación a tesis de grado:

- Alejandro Riveros trabajando en el uso de procesos y tecnologías de IN en la Dirección Provincial de Aeronáutica de la provincia de San Juan con el propósito de lograr una mejor gestión de los recursos, manteniendo las prioridades de la repartición.
- Mariano Alaniz en la aplicación de inteligencia de negocios en una pequeña empresa dedicada a la venta de productos informáticos, focalizada en mejorar la gestión de stock y la toma de decisiones en general.
- María Laura Molina desarrollando un modelo que permita relacionar datos y obtener conocimiento para generar perfiles de posibles estudiantes de carreras tecnológicas/informáticas, aplicando técnicas modernas de analítica de negocios, en base a los datos de aspirantes e ingresantes de diferentes años.
- Martín Manassero se encuentra iniciando su trabajo de tesis, relacionado a la propuesta de un sistema analítico y gestor de cartera de inversiones en criptomonedas.
- Leandro Drazic iniciando lineamientos sobre procesamiento de datos para IN en la nube, para su tesis de grado.

En cuanto a tesis de postgrado:

- La Lic. Cristina Vera se encuentra culminando el trabajo de su tesis de Maestría en Informática. La temática abordada refiere al uso (transaccional y de análisis) de gestores de bases de datos orientados a grafos y su comparación con los robustos y ampliamente conocidos relacionales.
- El Lic. Diego Checarelli, que se desempeña como profesional en la Dirección Provincial de Informática del gobierno de la provincia de San Juan, inició su trabajo de Tesis de Maestría en la temática de gestión y seguimiento del estado de avance de los proyectos estatales, utilizando técnicas propias de IN.

Se incorporaron en el segundo semestre del 2021 dos alumnos, Juan Capdevila y Martín Manassero, como alumno adscripto y alumno tesista, respectivamente.

5. BIBLIOGRAFÍA

- [1] H. Muñoz Hernández, R. C. Osorio Mass, and L. M. Zúñiga Pérez, “Inteligencia de los negocios. Clave del Éxito en la era de la información,” *CLIO América*, ISSN-e 2389-7848, ISSN 1909-941X, vol. 10, no. 20, pp. 194–211, 2016.
- [2] J. M. Rodríguez Parrilla, *Cómo hacer inteligente su negocio: business intelligence a su alcance*. Mexico: Grupo Editorial Patria., 2015.
- [3] R. Torres, A. Sidorova, and M. C. Jones, “Enabling firm performance through business intelligence and analytics: A dynamic capabilities perspective,” *Inf. Manag.*, vol. 55, no. 7, pp. 822–839, Nov. 2018.
- [4] J. C. V. Briano, *Sistemas de información gerencial: tecnologías para agregar valor a las organizaciones*, 1ª. Prentice Hall/Pearson Education, 2011.
- [5] K. C. Laudon and J. P. Laudon, *Sistemas de información gerencial*, 14ed ed. Pearson Educación, 2016.
- [6] J. Conesa Caralt and J. Curto Díaz, *Introducción al Business Intelligence*, 2da ed. Editorial UOC, 2011.
- [7] C. Loreti, “Aplicación de una solución de Business Intelligence que permita a miembros de pequeñas y medianas empresas gestionar la información para conseguir mejoras en las tomas de decisiones organizacionales,” U.Siglo XXI, 2016.
- [8] G. Bordoy, “Business Intelligence, la clave para que las pymes decidan con inteligencia,” *Cronista - PYME - Negocios - Business Intelligence*, 2018. [Online]. Available: <https://www.cronista.com/pyme/negocios/Business-Intelligence-la-clave-para-que-las-pymes-decidan-con-inteligencia-20181205-0005.html>.
- [9] “La inteligencia de negocio, clave para el desarrollo de las PYMES,” *Marketing Directo*, 15-Jan-2018. [Online]. Available: <https://www.marketingdirecto.com/digital-general/digital/la-inteligencia-negocio-clave-desarrollo-las-pymes>. [Accessed: 28-Sep-2021].
- [10] P. Clusellas, E. Martelli, and M. J. Martelo, *Un gobierno inteligente*, 1ra ed. CABA, Argentina: Boletín oficial, 2019.
- [11] “ARGENTINA: El 88% de las pymes digitalizó sus empresas a raíz del contexto pandemia,” *Business Trend*, 2021. [Online]. Available: <http://www.businesstrend.com.ar/argentina-el-88-de-las-pymes-digitalizo-sus-empresas-a-raiz-del-contexto-pandemia/>. [Accessed: 02-Mar-2022].
- [12] N. Mazón López, J. Pardillo Vela, and J. C. Trujillo Mondejar, *Diseño y explotación de almacenes de datos: Conceptos básicos de modelado multidimensional*. Editorial Club Universitario, 2011.
- [13] A. Vaisman and E. Zimányi, *Data Warehouse Systems - Design and Implementation*. Springer-Verlag Berlin Heidelberg, 2014.
- [14] J. Bhogal and I. Choksi, “Handling Big Data Using NoSQL,” in *Proceedings - IEEE 29th International Conference on Advanced Information Networking and Applications Workshops, WAINA 2015*, 2015.
- [15] R. Cattell, “Scalable SQL and NoSQL data stores,” *SIGMOD Rec.*, vol. 39, no. 4, 2010.
- [16] R. Angles and C. Gutierrez, “An introduction to graph data management,” *arXiv*. 2017. [Accessed: 27-Nov-2019].

Extracción y Explotación de Conocimiento para la Gestión en línea de datos en Ciencias del Mar

Marcos Zárate^{1,3}, Carlos Buckle¹, Renato Mazzanti^{1,2}, Mirtha Lewis^{3,4}, Gustavo Nuñez¹, Dario Ceballos³

¹ Laboratorio de Investigación en Informática, Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco (LINVI-FI-UNPSJB), Puerto Madryn, Argentina, +54 280-4883585 Int. 117.

² Unidad de Gestión de la Información, Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas (UGI-CENPAT-CONICET), Puerto Madryn, Argentina

³ Centro para el Estudio de Sistemas Marinos, Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas (CESIMAR-CENPAT-CONICET), Puerto Madryn, Argentina.

⁴ Centro de Investigaciones y Transferencia Golfo San Jorge, (CIT-GSJ-CONICET), Comodoro Rivadavia, Argentina.

zarate@cenpat-conicet.gob.ar, cbuckle@unpata.edu.ar, renato@cenpat-conicet.gob.ar, mirtha@cenpat-conicet.gob.ar, guscostaf@gmail.com, disenodc@gmail.com

Resumen

El objetivo general de esta investigación es estudiar y desarrollar grafos de conocimiento para la gestión integrada y visualización de datos en línea de ciencias del mar a través de tecnologías Big Data y datos provenientes de campañas oceanográficas, de repositorios de Biodiversidad marina y de datos ambientales. En una primera etapa será acotado al golfo San Jorge (GSJ) dado que se cuenta con datos provenientes de las campañas realizadas por el grupo de trabajo GSJ, perteneciente a la iniciativa Pampa Azul y luego puede ser escalable a otros espacios marinos donde se cuenta con información de campañas oceanográficas así como estaciones fijas con sensores ambientales remotos. El impacto esperado de esta investigación será el de permitir una gestión confiable de los datos y una explotación adecuada en línea, para garantizar la preservación y el acceso a nuestros activos nacionales de investigación multidisciplinar. El proyecto se realiza en conjunto entre LINVI-FI-UNPSJB y CESIMAR-CENPAT-CONICET. En el proyecto participan docentes investigadores de la carrera de Licenciatura en Informática, investigadores, un becario post-doctoral y un

becario CONICET y un estudiante avanzado aspirante a beca EVC-CIN.

Palabras clave: Grafos de Conocimiento, Datos Abiertos Enlazados, Datos Oceanográficos, Grandes Volúmenes de Datos.

Contexto

Esta propuesta avanzará sobre resultados y líneas de investigación del proyecto precedente *UNPSJB-PI 1562 - Plataforma de Datos Abiertos Enlazados para la Gestión y Visualización de Datos Primarios de Ciencias del Mar*, en el cual se identificaron ventajas en los Datos Abiertos Enlazados (Linked Open Data) [1] como camino estándar hacia la integración de datos abiertos [2] heterogéneos de diferentes dominios, bajo vocabularios comunes y con posibilidades de razonamiento por partes de agentes de software. Este trabajo se focalizará específicamente en las estrategias y técnicas de visualización de datos científicos relacionados con las ciencias del mar para poder ser explotados de manera adecuada por los investigadores.

Estos proyectos tienen una concepción interdisciplinaria y por ello el grupo de trabajo incluye: investigadores de las ciencias de la

computación de diferentes áreas e investigadores de las ciencias biológicas, expertos en el dominio de aplicación.

Este proyecto está avalado por el Consejo Directivo de la Facultad de Ingeniería de la UNPSJB y será financiado por la Secretaría de Ciencia y Técnica de la UNPSJB para llevar a cabo durante el periodo 2022-2023.

1. Introducción

En el contexto de la iniciativa Pampa Azul [3] relanzada en julio de 2020, la gestión y modelado de datos en línea es considerada una disciplina fundamental para abordar la complejidad y el alcance de las temáticas que exigen una aproximación interdisciplinaria y una proyección amplia en el uso de la información. Por lo tanto, es necesario desarrollar sistemas capaces de gestionar su integración y su comunicación, tanto para un aprovechamiento integral y secundario de los grupos e instituciones participantes como para usuarios externos que requieran de información.

La relevancia de la propuesta se identifica en diferentes niveles: **(1)** a nivel tecnológico, la creación de una plataforma que permita gestionar datos de ciencias del mar, permitirá resolver los problemas de integración de datos heterogéneos bajo un mismo vocabulario y disponerlos para ser explotados mediante búsquedas semánticas y consultas basadas en lenguaje natural. Adicionalmente, el desarrollo de visualizaciones de datos relacionados a las ciencias del mar (Data Visualization) [3] facilitará su interpretación y comparación. **(2)** a nivel de política científica para la región y el país, se logrará un prototipo de servicio estandarizado para repositorios de datos marinos del Atlántico Sur Occidental, con información unificada provenientes de diferentes plataformas de muestreo (buques oceanográficos, ROVs, animales instrumentados, etc.), que podrá ser reusada y explotada para la generación de nuevos conocimientos. **(3)** a nivel socioeconómico, las posibilidades de reuso de datos permitirán reducir los altos costos de las campañas de

investigación en el mar y aportará al uso democrático y eficiente de datos científicos. Finalmente **(4)** a nivel de formación de recursos humanos del proyecto participa un becario doctoral CONICET de la disciplina Ciencias de la Computación e Informática y un aspirante a becas EVC-CIN.

2. Motivación

En Julio de 2020 se relanzó la iniciativa Pampa Azul [4], destinada a promover el conocimiento científico, el desarrollo tecnológico y la innovación productiva en el Atlántico Sur, con el fin de crear una cultura del mar en la sociedad argentina, fomentar el uso sostenible de los bienes naturales marinos y fortalecer el crecimiento de la industria nacional asociada.

Sin embargo, el manejo en línea de los datos generados en el primer lanzamiento de Pampa Azul (año 2014) dejó en claro que la gestión y modelado de los datos en línea necesitaba ser planificado, resguardado y compartido para que los productos que se generen con ellos puedan ser utilizados por la comunidad científica para una adecuada comprensión del funcionamiento de nuestros espacios marinos. Más aún, no se tuvo en cuenta la integración de estos datos con bases federadas de alcance mundial, por lo cual se hace muy dificultoso el aprovechamiento científico integral de los mismos.

3. Líneas de Investigación y Desarrollo

Este proyecto desarrolla como principal línea de investigación, el Modelado Conceptual en la Web Semántica [5] y la construcción de grafos de conocimiento oceanográfico [6, 7] mediante datos enlazados [8, 9] para la integración de datos científicos y su explotación [10, 11]. Pero además, será necesario abordar el tratamiento de grandes volúmenes de datos oceanográficos y meteorológicos [12], caracterizados por las 5V (Volumen, Velocidad, Variedad, Verosimilitud y Valor) de Big-data [13, 14] y teorías, técnicas y herramientas para la

visualización de datos científicos oceanográficos [15, 16].

4. Resultados esperados

El objetivo general del proyecto es estudiar y desarrollar grafos de conocimiento para la gestión integrada y visualización de datos en línea de ciencias del mar a través de tecnologías Big Data integrando datos provenientes de campañas oceanográficas, de repositorios de Biodiversidad marina y de datos ambientales. En una primera etapa será acotado al golfo San Jorge (GSJ) dado que se cuenta con datos provenientes de las campañas realizadas por el grupo de trabajo GSJ, perteneciente a la iniciativa Pampa Azul y luego puede ser escalable a otros espacios marinos donde se cuente con información de campañas oceanográficas así como estaciones fijas con sensores ambientales remotos. El impacto esperado de esta investigación será el de permitir una gestión confiable de los datos en línea, para garantizar la preservación y el acceso a nuestros activos nacionales de investigación multidisciplinar.

Se esperan como resultados:

1. Extender el modelo conceptual BiGe-Onto [17] y formalizar el grafo de conocimiento OceanGraph [18] desarrollado durante el proyecto UNPSJB-PI 1562, incluyendo nuevos registros que agregan complejidad a la información.
2. Estudio y desarrollo de una plataforma de visualización en línea para proveer facilidades de analíticos visuales y permitir consultas y análisis interactivos de la información proveniente de los grafos de conocimiento.
3. Estudiar e implementar sistemas de razonamiento que permitan la explotación de conocimiento implícito en el grafo.
4. Implementar el plan de gestión de datos que describe la planificación, organización y previsión de la colección digital de los datos recolectados, reutilizados o procesados. Para su construcción se toma como guía el Plan de Gestión de Datos CONICET.

5. Formación de recursos humanos

En este proyecto participa un docente de UNPSJB-Puerto Madryn, doctor en Ciencias de la Computación y actual becario posdoctoral CONICET focalizado en el desarrollo de Grafos de Conocimiento para la gestión integrada de datos científicos multidisciplinares de ciencias oceanográficas. También participa un becario doctoral CONICET perteneciente a CESIMAR-CONICET que investiga analíticos visuales para datos enlazados en Ciencias del Mar. Sus directores son investigadores de la UNPSJB, de la Universidad Nacional del Sur y del CESIMAR-CENPAT-CONICET, quienes también integran el equipo de trabajo en calidad de asesores y expertos de dominio.

Además, el equipo de trabajo incluye a tres docentes del Departamento de Informática de la Facultad de Ingeniería de la UNPSJB-Puerto Madryn, que desarrollan líneas de investigación referidas a la Gestión de Información Científica y Tecnológica.

Referencias

- [1] Tom Heath and Christian Bizer. Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1(1):1–136, 2011.
- [2] Carol Tenopir, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. Data sharing by scientists: Practices and perceptions. *PLoS ONE*, 6, 06/2011 2011.
- [3] Schlitzer, R. (2002). Interactive analysis and visualization of geoscience data with Ocean Data View. *Computers & geosciences*, 28(10), 1211-1218.
- [4] Pampa Azul. <https://www.argentina.gob.ar/noticias/se-relanzo-la-iniciativa-pampa-azul>, 2020. [Online; accessed 22-Feb-2022].

- [5] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific American*, 284(5):34–43, May 2001.
- [6] Marcos Zárate, Pablo Rosales, Pablo Fillottrani, Claudio Delrieux, and Mirtha Lewis. Oceanographic data management: Towards the publishing of pampa azul oceanographic campaigns as linked data. In *Proceedings of the 12th Alberto Mendelzon International Workshop on Foundations of Data Management (AMW 2018)*, 2018.
- [7] Marcos Zárate, Pablo Rosales, Germán Braun, Mirtha Lewis, Pablo Rubén Fillottrani, and Claudio Delrieux. OceanGraph: Some initial steps toward a oceanographic knowledge graph. In *Iberoamerican Knowledge Graphs and Semantic Web Conference*, pages 33–40. Springer, 2019.
- [8] Krzysztof Janowicz, Pascal Hitzler, Benjamin Adams, Dave Kolas, II Vardeman, et al. Five stars of linked data vocabulary use. *Semantic Web*, 5(3):173–176, 2014.
- [9] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked data: The story so far. In *Semantic services, interoperability and web applications: emerging concepts*, pages 205–227. IGI Global, 2011.
- [10] Judith A Blake and Carol J Bult. Beyond the data deluge: data integration and bioontologies. *Journal of biomedical informatics*, 39(3):314–320, 2006.
- [11] Giuseppe Andronico, Valeria Ardizzone, Roberto Barbera, Bruce Becker, Riccardo Bruno, Antonio Calanducci, Diego Carvalho, Leandro Ciuffo, Marco Fargetta, Emidio Giorgio, Giuseppe Rocca, Alberto Masoni, Marco Paganoni, Federico Ruggieri, and Diego Scardaci. e-infrastructures for e-science: A global view. *Journal of Grid Computing*, 9:155–184, 06 2011.
- [12] Adam Leadbetter, Robert Arko, Cynthia Chandler, Adam Shepherd, and Roy Lowry. Linked Data An Oceanographic Perspective. *The Journal of ocean Technology*, 8(3), 2013.
- [13] Mark A Beyer and Douglas Laney. The importance of ‘big data’: a definition. *Stamford, CT: Gartner*, pages 2014–2018, 2012.
- [14] Nikos Bikakis and Timos Sellis. Exploration and visualization in the web of big linked data: A survey of the state of the art. *arXiv preprint arXiv:1601.08059*, 2016.
- [15] Tanu Malik and Ian Foster. Addressing data access needs of the long-tail distribution of geoscientists. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 5348–5351. IEEE, 2012.
- [16] Peter Fox and James Hendler. Changing the equation on scientific data visualization. *Science*, 331(6018):705–708, 2011.
- [17] Zárate, M., Braun, G., Fillottrani, P., Delrieux, C., & Lewis, M. (2020). BiGe-Onto: an ontology-based system for managing biodiversity and biogeography data. *Applied Ontology*, 15(4), 411-437.
- [18] Zárate, M., Rosales, P., Braun, G., Lewis, M., Fillottrani, P. R., & Delrieux, C. (2019). OceanGraph: Some initial steps toward a oceanographic knowledge graph. In *Knowledge Graphs and Semantic Web* (pp. 33–40). Springer International Publishing. https://doi.org/10.1007/978-3-030-21395-4_3

Algoritmos de detección de Outliers

María Fernanda Cuadrado, Adrián Alfredo De Armas, Alejandro Luis Foglino

Universidad Argentina de la Empresa (UADE)
{mecuadrado, adearmas, afoglino}@uade.edu.ar

RESUMEN

Esta investigación aborda la problemática de la detección de outliers en grandes bases de datos mediante distintos tipos de algoritmos. Los tipos de algoritmos se clasifican profundidad, densidad, desvío, ángulos y distancia. Se implementarán distintas versiones de algoritmos por cada tipo y se paralelizará su ejecución buscando mejorar la eficiencia a medida que el conjunto de datos utilizado crezca.

Palabras Clave: outliers, densidad, distancia, ángulos, grandes bases de datos.

CONTEXTO

Una de las definiciones más citadas en la bibliografía respecto a lo que es un outlier es la enunciada por David Hawkins en sus monografías sobre estadística y probabilidad aplicadas del año 1980: “Un outlier es una observación que se desvía tanto de otras observaciones que despierta la sospecha de haber sido generado por un mecanismo diferente” [4] [Hawkings, 1980].

Existen diversos métodos para la detección de outliers que van desde la determinación de los modelos estadísticos y probabilísticos de los datos hasta métodos basados en [2] [AGGARWAL, 2013]:

1. Profundidad / Densidad
2. Desvío
3. Ángulos
4. Distancia

El objetivo de la presente investigación es la implementación de al menos un algoritmo para los métodos mencionados con la particularidad de optimizar las implementaciones para ser

usadas en grandes bases de datos, y el aprovechamiento de las computadoras multi núcleo implementando versiones del algoritmo que aprovechen la paralelización de operaciones.

Para los métodos de profundidad, desvío, ángulos y distancia se analizarán los distintos algoritmos disponibles, se los implementará y se realizarán comparaciones de funcionamiento y rendimiento en distintos conjuntos de datos.

Este trabajo permitirá conocer el estado situación actual de los algoritmos de detección de outliers. Se expondrán las decisiones de arquitectura y desarrollo que se toman a lo largo de este proyecto para permitir críticas que alimenten una discusión que habilite la mejora de los algoritmos desarrollados.

1. INTRODUCCIÓN

Mantener la calidad de los datos es un activo clave en los procesos de decisión de las organizaciones, es una actividad que requiere controles reiterados en distintas instancias desde el ingreso, almacenamiento y uso.

Es cada vez más generalizado el uso de los datawarehouse en las empresas. Los mismos almacenan todos los datos relevantes para la organización o, incluso aún, aquellos que tienen potencial de resultar relevantes en el futuro, aunque al momento de decidir su almacenamiento esto no se tenga tan claro.

El datawarehouse se utiliza como un repositorio de datos para luego poder procesarlo y brindarle información útil al usuario.

Las decisiones que se toman dentro de la organización dependen de la interpretación de la información proveniente del procesamiento de

los datos contenidos en el datawarehouse, por este motivo, es de suma importancia medir la calidad de los datos con la que se cuenta ya que del procesamiento de estos se obtendrán las salidas que serán luego evaluadas.

Con los datos contenidos en el repositorio de datos (datawarehouse) se pueden aplicar distintas técnicas de datamining. *Datamining* se denomina al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos. Estos patrones pueden ser interpretados como un resumen de los datos de entrada. La real tarea de la minería de datos es el análisis automático de grandes bases de datos para la extracción de patrones de utilidad, hasta ahora desconocidos como las dependencias (grupo de datos dependiendo funcionalmente de otro), grupos de datos (los datos se agrupan formando distintos grupos de interés), y anomalías (datos que son inconsistentes con el grueso de los datos, también conocidos como outliers).

Todos los sistemas de la organización deberían brindar datos a este repositorio como actividad permanente, teniendo como consecuencia que el crecimiento del datawarehouse sea constante en el tiempo.

El crecimiento de los datawarehouse nos pone frente al desafío de poder procesar los datos para convertirlos en información, y sobre todo poder medir la calidad de los datos contenidos en el mismo.

Ésta investigación se concentrará en trabajar sobre grandes bases de datos (datawarehouse o sistemas de producción) y para demarcar un límite, consideraremos “grande” a una base de datos cuyo análisis requiera del procesamiento de, al menos, 10 millones de filas.

A medida que el volumen de datos crece, las relaciones aumentan en cantidad, entrecruzamiento y acoplamiento, sumado al paso del tiempo, la desactualización se incrementa afectando el valor de los datos.

Una anomalía representa una irregularidad en los datos que contiene la base de datos que puede ser de distinta naturaleza:

- *Inconsistencia de relación*: Son aquellas inconsistencias que no respetan las reglas de integridad referencial estipuladas en la base de datos. Por ejemplo, si hubiera una base de datos con una tabla de tipos de documento (1-DNI, 2-Libreta cívica y 3-Libreta de enrolamiento) y una tabla de clientes con un campo de “tipo de documento” relacionado con la tabla de tipos de documento, una inconsistencia de relación sería encontrar en la tabla de clientes algún registro con un 4 como valor del campo tipo de documento. Si bien, las bases de datos hacen el control de integridad referencial al insertar un registro, cuando se hace un vuelco de grandes volúmenes de datos provenientes de un sistema, este control (el de integridad referencial) suele deshabilitarse para poder lograr que el proceso de volcado pueda ser llevado a cabo.
- *Inconsistencia de comportamiento*: Son aquellas inconsistencias conocidas también como outliers. Los outliers representan “datos que son significativamente diferentes a otros datos de la colección o un elemento que parece implicar un patrón que es inconsistente con el grueso de la evidencia de datos”. [3] [mathematics dictionary, 2007]

Muchos algoritmos de datamining buscan minimizar la influencia de los outliers o directamente los eliminan. Esto podría resultar en la pérdida de importante información que se encuentra oculta. La detección de outliers puede ser de particular interés ya que uno de los usos que permite es la identificación de actividad fraudulenta, accesos no permitidos y vuelco de datos inconsistentes, entre otros. La técnica de detección de outliers encuentra aplicación en detección de fraudes con tarjetas de crédito, análisis de robustez de redes, detección de

intrusiones en redes, aplicaciones financieras y de marketing [1] [Abu Bakar y otros, 2006].

La calidad de los datos que se encuentran presentes en una base de datos no puede ser evaluada sino es a partir de herramientas con algoritmos como los que se desarrollarán, y ahí es donde radica la importancia de explorar, medir y comparar los resultados de la aplicación de las distintas técnicas, esperando encontrar variedad de resultados (detección de datos anómalos), de rendimiento y consumo de recursos.

El aporte esencial de esta propuesta es que todo el análisis se realice de forma genérica, sin buscar anomalías específicas influenciadas por el área de aplicación.

Los algoritmos disponibles para detección de outliers que implementa cualquier técnica no tienen en cuenta la cantidad de recursos disponibles en el sistema (como cantidad de memoria, disponibilidad del procesador, cantidad de disco disponible para tablas temporales, etc.) y, tratándose de grandes bases de datos, donde cualquier proceso que se quiera realizar, por ejemplo una simple sumatoria) puede requerir de una cantidad importante de tiempo de procesador y memoria, lo cual plantea un desafío a la hora de adaptar los algoritmos para poder abarcar dicho volumen de datos y aún mantener su funcionalidad.

2. LINEAS DE INVESTIGACION Y DESARROLLO

El presente proyecto se encuadra dentro de la ingeniería de software empírica. Se propone la implementación de distintos algoritmos de detección de outliers para poder relevar sus fortalezas y limitaciones con el fin de maximizar sus fortalezas y acotar sus limitaciones cuando el algoritmo sea ejecutado en un conjunto de datos cada vez más grande. Esta investigación remota el trabajo realizado en el año 2015 en el proyecto de investigación sobre detección de outliers con algoritmos de distancia donde se implementaron distintas versiones del algoritmo FindAllOutsM [5][De Armas y otros, 2015]

3. RESULTADOS OBTENIDOS/ESPERADOS

Se espera poder medir el rendimiento de los distintos tipos de algoritmos para diferentes características de outliers y clasificarlos según conveniencia de aplicación según las características propias del conjunto de datos utilizado (distribución y tamaño). Se conformará una guía de selección del tipo de algoritmo a utilizar en base al conjunto de dato a procesar que, junto con el código del algoritmo implementado, permita al lector la selección e implementación de un algoritmo de detección de outliers según las características de su caso particular. Se producirá un software que permita ejecutar todo lo producido en el ámbito de este proyecto de investigación.

4. FORMACIÓN DE RECURSOS HUMANOS

Los participantes de esta investigación (alumnos de grado y maestría) serán capacitados en la implementación de técnicas matemáticas/estadísticas en distintos lenguajes de programación. Realizaran la verificación de los algoritmos implementados y la comparación entre las distintas implementaciones para determinar puntos de referencia respecto del mejor algoritmo para cada escenario.

5. BIBLIOGRAFÍA

- [1] Abu Bakar Zuriana, Mohemad Rosmayati, Ahmad Akbar. A Comparative Study for Outlier Detection Techniques in Data Mining. Department of Computer Science Faculty of Science and Technology, University College of Science and Technology 21030 Kuala Terengganu, Malaysia. 2006
- [2] AGGARWAL C. Outliers Analysis. Springer, IBM T.J. Watson Research Center, Yorktown Heights, New Work, USA. 2013.

- [3] Edu2000 America Inc., página web. Mathematics Dictionary, 1995-2007. Consulta realizada el 16 de mayo del 2013.
<http://www.mathematicsdictionary.com/english/vmd/full/o/outlier.htm>
- [4] HAWKINS D.. Identification of Outliers (Monographs on Statistics and Applied Probability) vol 3. Chapman and Hall, London. 1980.
- [5] De Armas Adrián y otros. Detección de outliers en grandes bases de datos mediante aproximación basada en celdas. 2015. Consulta realizada el 2 de marzo de 2022.
http://sedici.unlp.edu.ar/bitstream/handle/10915/52173/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y

Chequeo de Consistencia de Procedimientos Almacenados y Triggers

Joaquín Gardiola, Fabio Zorzan, Mariana Frutos, Marcela Daniele, Ariel Arsaute

Departamento de Computación

Facultad de Ciencias Exactas, Físico-Químicas y Naturales - Universidad Nacional de Río Cuarto

joa.gardiola@gmail.com, {fzorzan, mfrutos, marcela, aarsaute }@dc.exa.unrc.edu.ar

CONTEXTO

La línea de investigación presentada en este trabajo se desarrolla en el marco del proyecto “Ingeniería de Software: Evaluación de la calidad de sistemas de software y la mejora continua de los procesos de desarrollo”, presentado en el Programa de Convocatoria PPI 2019 de la Secretaría de Ciencia y Técnica de la Universidad Nacional de Río Cuarto.

RESUMEN

Esta línea de investigación pretende contribuir con un aporte al desarrollo de aplicaciones que utilizan triggers y procedimientos almacenados en Bases de datos Relacionales, en particular PostgreSQL y MySQL, proponiendo el desarrollo de una herramienta que permita realizar chequeos semánticos adicionales a las instrucciones SQL incluidas en los códigos implementados en triggers y procedimientos Almacenados, permitiendo de esta manera proveer una solución que facilite detectar de manera temprana, problemas semánticos en las instrucciones SQL incluidas en estos mecanismos de programación de bases de datos relacionales.

Palabras Clave: Triggers, Procedimientos almacenados, chequeo semántico, SQL, MySQL, PostgreSQL

1. INTRODUCCIÓN

En la actualidad las bases de datos relacionales son herramientas ampliamente utilizadas en el desarrollo de sistemas de información[1]. Las bases de datos relacionales proveen mecanismos

para programar dentro de la base de datos, estos mecanismos se denominan procedimientos almacenados y triggers[2], estas características se pueden encontrar en los motores de bases de datos MySQL[3] y PostgreSQL[4].

Cuando una base de datos PostgreSQL y MySQL cuenta con triggers y procedimientos almacenados, cuando se crean, se chequean sintáctica y semánticamente, pero hay ciertos chequeos semánticos que no se realizan, por ejemplo, se puede escribir como parte del código de un trigger una consulta SQL sobre tablas que no existen.

El crecimiento sostenido de la industria del software y su constante expansión a nuevos ámbitos, genera un estado de alerta y constante revisión de las metodologías de desarrollo de software utilizadas, como así también de los métodos de trabajo, la adopción de nuevas herramientas para automatizar y mejorar las actividades de gestión e ingeniería, como la planificación, el diseño y la implementación, involucradas en el proceso de producción de software. A pesar de las investigaciones y estudios realizados, existen aún numerosos casos en los que no se ha podido evitar que los errores de software lleguen a etapas de producción. El proyecto marco propone mejorar la calidad de los sistemas de software que se desarrollan, continuando con la investigación y el estudio de técnicas y procesos de desarrollo de software con diferentes enfoques.

El objetivo principal de esta línea es el desarrollo de una herramienta que pretende mejorar la calidad del software, en particular

software que utiliza una base de datos relacional, detectando en una etapa temprana del desarrollo, antes de llegar al entorno de producción, posibles inconsistencias del código SQL incorporado en procedimientos almacenados y triggers.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

El trabajo está planificado para desarrollarse en tres etapas que se detallan a continuación:

2.1. PRIMERA ETAPA

En una primera etapa (ya concluida), se realizó un estudio del estado del arte y se analizaron las herramientas existentes que abordan, en parte, el problema planteado. Luego de hacer una revisión en portales científicos, distintos buscadores, documentos de investigación, foros, etc., se pudo observar que hasta la actualidad las formas de solucionar el problema de chequeo de consistencia en procedimientos almacenados y triggers en MySQL y PostgreSQL son algo rústicas, y no atacan al problema de una manera específica, como sería la creación de una herramienta de software que permita realizar chequeos semánticos de consistencia en el código de procedimientos almacenados y triggers en estas base de datos.

Una de las formas que se encontró como más “habitual” para resolver dicho problema, pero poco práctico, fue seguir la siguiente secuencia:

1. Realizar una copia de la Base de Datos.
2. Crear un script (secuencia de comandos que se utiliza para automatizar tareas repetitivas) con todos los procedimientos almacenados y triggers.
3. Realizar un “drop” (eliminación) de todos los procedimientos almacenados y triggers.
4. Por último, correr el script para recrearlos.

Siguiendo el procedimiento anterior, los procedimientos y triggers que no se puedan recrear arrojarán un error. Este método, si bien en parte da respuesta al problema, no detecta todos los problemas que se pueden detectar con un análisis a fondo usando una herramienta a tal fin.

Por ejemplo, hay triggers que no arrojan errores cuando se crean, pero sí cuando se ejecutan. Muchos de estos errores se podrían detectar de manera estática. La idea del presente trabajo es no tener que llegar a detectar el problema en tiempo de ejecución, sino permitir su detección en una etapa más temprana del desarrollo con un chequeo semántico estático.

2.2. SEGUNDA ETAPA

En esta etapa se definieron que tipos de chequeos específicos de bases de datos se implementarán en una primera instancia del trabajo. Además se realizó un estudio y evaluación de diferentes herramientas de análisis léxico – sintáctico y semántico, y del lenguaje de programación para el desarrollo. De esta forma se evaluó cuáles son las que mejor se adaptaban a los requisitos de la aplicación a desarrollar.

El lenguaje seleccionado fue Java, que además de contar con herramientas de análisis léxico – sintáctico y semántico, como son las librerías JFlex [5] para análisis léxico y Java_Cup [6] y Java_Cup-11a [6] para análisis sintáctico, provee mecanismos de fácil conexión a bases de datos PostgreSQL y MySQL. Cabe aclarar que la primera versión de la herramienta está enfocada en el motor PostgreSQL.

2.3. TERCERA ETAPA

Durante esta etapa se comenzó con el proceso para la construcción de la herramienta, la cual dispone de una fase de análisis donde se interpretan las entradas y se generan estructuras intermedias.

Esta fase estará estructurada de la siguiente forma:

- Análisis Léxico
- Análisis Sintáctico
- Análisis Semántico

A continuación, se describen cada una de estas fases:

2.3.1. ANÁLISIS LÉXICO

El analizador léxico, esquematizado en la Fig. 1, cumple la función de scanner, lee los caracteres

uno a uno desde la entrada y va formando grupos de caracteres con alguna relación entre sí (tokens), que constituirán la entrada para la siguiente etapa (análisis sintáctico). Cada token representa una secuencia de caracteres que son tratados como una única entidad. Por ejemplo, en SQL un token es la palabra reservada SELECT.

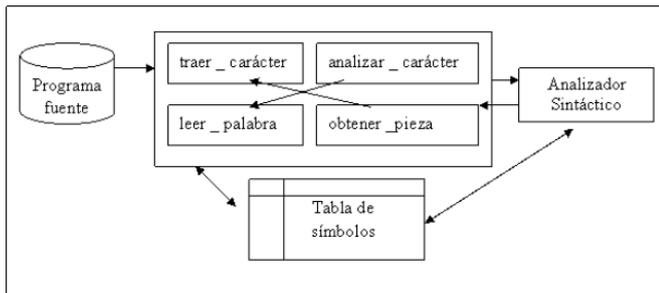


Fig.1: Esquema analizador Lexicográfico

Las funciones que realizará el analizador implementado en JFLEX son:

- Leer los caracteres de entrada del programa fuente (código fuente de Triggers y Procedimientos Almacenados).
- Generar la secuencia de componentes léxicos.
- Eliminar comentarios, espacios en blanco, y todo lo que esté excluido de la sintaxis del lenguaje.
- Reconocer identificadores, palabras reservadas del lenguaje, y tratarlos correctamente con respecto a la tabla de símbolos (solo en los casos que deben tratarse con dicha tabla).
- Avisar de errores léxicos. Por ejemplo, si @ no pertenece al lenguaje, avisar de un error.

Para llevar a cabo esta fase, se debe especificar las reglas léxicas. A continuación se muestra un fragmento de las reglas especificadas en el analizador lexicográfico implementado en JFLEX:

```

import static codigo.Tokens.*;
%%
%class Lexer
%type Tokens
%ignorecase

L=[a-zA-Z_]+
D=[0-9]+
espacio=[ ,\t,\r]+
%{
    public String lexeme;
%}
%%

( "," ) {lexeme = yytext(); return COMA;}

...

/* Palabra reservada Select*/
"select" {lexeme=yytext(); return SELECT;}
/* Palabra reservada Distinct */
( distinct _ ) {lexeme=yytext(); return
DISTINCT;}
/* Palabra reservada From */
( from _ ) {lexeme=yytext(); return FROM;}
/* Palabra reservada Where */
( where _ ) {lexeme=yytext(); return WHERE;}

...
  
```

2.3.2. ANÁLISIS SINTÁCTICO

El análisis sintáctico, esquematizado en la fig. 2, también llamado parser, se encargará de chequear el texto de entrada, suministrado por el analizador léxico, en base a una gramática dada (gramática del lenguaje SQL). En caso que el texto corresponda a una sentencia válida, suministrará el árbol sintáctico que la reconoce. Además de determinar si la gramática establecida por el lenguaje se respeta, el analizador sintáctico a implementar en Java_Cup tendrá las siguientes funciones:

- Acceder a la tabla de símbolos para armar el árbol sintáctico.
- Chequear los tipos.
- Generar código intermedio.
- Generar errores cuando se producen.

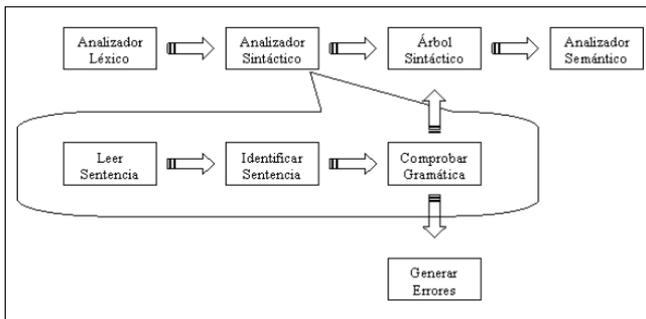


Fig.2: Esquema analizador Sintáctico

A continuación se muestra un fragmento de la especificación del analizador sintáctico implementado en JAVA_CUP:

```

import java_cup.runtime.Symbol;

parser code
{
    private Symbol s;

    public void syntax_error(Symbol s){
        this.s = s;
    }

    public Symbol getS(){
        return this.s;
    }
};
--
terminal SELECT, DISTINCT, FROM, WHERE,
MAYOR, DESC, ASC, GROUP, HAVING, ORDER,
    ID, NUM, P_COMA, COMA, OR, AND, EQ,
MENOR, LE, GE, NE, LIKE;
non terminal S, ST1, ST2, ST3, ST4, ST5,
ST6, attributeList, tableList, COND, E, F;

/* Precedencias */
precedence left ID, NUM;
precedence left OR, AND;
precedence left GROUP, HAVING, ORDER,
P_COMA, COMA;

start with S;
S ::= ST1 P_COMA;
ST1 ::= SELECT attributeList FROM tableList ST2
    | SELECT DISTINCT attributeList FROM
    tableList ST2;
ST2 ::= WHERE COND ST3
    | ST3;
ST3 ::= GROUP attributeList ST4
    | ST4;

```

2.3.3. ANALIZADOR SEMÁNTICO

El analizador semántico revisa que las operaciones sean realizables al determinar que los tipos de las entidades sean correctos y que los operadores son capaces de manipularlos.

Algunas de las funciones más importantes, relacionadas específicamente al chequeo del código SQL incorporado en triggers y procedimientos almacenados, que se deberán realizar son:

- Comprobación de tipos en expresiones de filtros: La aplicación de los operadores y operandos deben ser compatibles (puede ser constantes, funciones, columnas de tablas, etc.).
- Comprobación de existencia de tablas y columnas en la base de datos.

Para implementar estos chequeos semánticos la herramienta deberá acceder al catálogo de la base de datos donde están definidos los procedimientos almacenados y triggers.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados de la investigación aportarán a la mejora del desarrollo de aplicaciones que utilizan Bases de datos Relacionales, y en particular, sus mecanismos de programación como lo son procedimientos almacenados y triggers, esto se logrará con el desarrollo de una herramienta de chequeos semánticos, adicionales a los realizados por los motores de bases de datos, mediante la utilización de herramientas para el desarrollo de parsers de lenguajes de programación.

El beneficio de esta herramienta se verá fundamentalmente en el desarrollo de aplicaciones que utilizan procedimientos almacenados y triggers en bases de datos relacionales, en particular PostgreSQL, detectando errores semánticos en una etapa temprana y antes que se produzcan en el entorno de producción.

En síntesis, los principales objetivos planteados en esta investigación son:

Específicos:

- ✓ Implementar un chequeador semántico de código SQL incorporado en procedimientos almacenados y triggers.
- ✓ Integrar la solución a desarrollar a herramientas oficiales de administración de bases de datos.

De Formación de los estudiante

- ✓ Potenciar el desarrollo de metodologías de trabajo científicas.
- ✓ Profundizar los conocimientos de ambientes de desarrollo y testing de aplicaciones.
- ✓ Desarrollar capacidades para la redacción de artículos científicos y documentación técnica.
- ✓ Profundizar la experiencia en el uso de bases de datos relacionales, en particular PostgreSQL.
- ✓ Adquirir experiencia en el uso de chequeadores sintácticos y semánticos de lenguajes de programación.

4. FORMACION DE RECURSOS HUMANOS

Los estudios realizados en esta línea de investigación sirven como marco para la elaboración de trabajos finales de grado. En este punto, dos estudiantes de la carrera Licenciatura en Ciencias de la Computación han iniciado su trabajo final que consiste en el desarrollo de la herramienta de cheque semántico presentada en este trabajo.

Los temas abordados en esta línea de investigación brindan un fuerte aporte al proceso de perfeccionamiento continuo de los autores de este trabajo, que se desempeñan como docentes de las carreras de computación que se dictan en la Universidad Nacional de Río Cuarto y participan en asignaturas relacionadas a dichos temas.

5. BIBLIOGRAFIA

- [1] Database System Concepts. 7th Edition, Edition Silberschatz, Korth, Sudarshan. McGraw Hill Company, 2019.
- [2] Fundamentals of Database Systems. Elmasri, Navathe. 5th Edition Addison Wesley, 2006.
- [3] Manual Oficial del motor de Bases de Datos MySQL, <http://downloads.mysql.com/docs/refman-8.0-en.a4.pdf>. Ultimo acceso 04/03/2022.
- [4] Manual Oficial del motor de Bases de Datos PostgreSQL, <https://www.postgresql.org/files/documentation/pdf/13/postgresql-13-A4.pdf>. Ultimo acceso 01/03/2022.
- [5] Manual oficial de JFlex <https://www.jflex.de/>. Ultimo acceso 23/02/2022.
- [6] LALR Generator for JAVA <http://www2.cs.tum.edu/projects/cup/>. Ultimo acceso 24/02/2022.

Procesamiento inteligente de la información. Aplicaciones en bioinformática, trayectorias vehiculares, mantenimiento preventivo industrial y sistemas embebidos

W. Hasperué¹, C. Estrebou¹, G. Camele^{1,3}, P. López², M. Peña², G. Reyes Zambrano⁴
L. Lanzarini¹, A. Fernandez Bariviera⁵, M. Cerrada⁶

¹ Instituto de Investigación en Informática LIDI*, Facultad de Informática, UNLP, La Plata, Argentina

² Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina

³ Becario postgrado UNLP

⁴ Facultad de Ciencias Físicas y Matemáticas, Universidad de Guayaquil, Guayaquil, Ecuador

⁵ Dpto. de Economía, Universitat Rovira i Virgili, Reus, España

⁶ GIDTEC, Universidad Politécnica Salesiana, Cuenca, Ecuador

* Centro asociado de la Comisión de Investigaciones Científicas de la Pcia. De Bs. As. (CIC)

{whasperue, cesarest, gcamele, pdlopez, laural}@lidi.info.unlp.edu.ar, mario.pena@info.unlp.edu.ar, gary.reyesz@ug.edu.ec, martinezvictor@hotmail.com, aurelio.fernandez@urv.net, mcerrada@ups.edu.ec

CONTEXTO

Esta presentación corresponde a las tareas de investigación que se llevan a cabo en el III-LIDI en el marco del proyecto “Sistemas inteligentes. Aplicaciones en reconocimiento de patrones, minería de datos y big data” perteneciente al Programa de Incentivos (2018-2022).

RESUMEN

Esta línea de investigación se centra en el estudio y desarrollo de Sistemas Inteligentes para la resolución de problemas de Big Data y Minería de Datos utilizando técnicas de Aprendizaje Automático. Los sistemas desarrollados se aplican particularmente al procesamiento de grandes volúmenes de información y al procesamiento de flujo de datos.

Las investigaciones correspondientes al procesamiento de datos masivos están enfocadas en dos temas: el estudio y desarrollo de técnicas de selección de características y el diseño de librerías que faciliten el procesamiento masivo de datos. En lo referido a selección de características, el foco está puesto en la reducción de los tiempos de cómputo. La optimización puede realizarse tanto en la mejora de la ejecución en un entorno distribuido, como en la propuesta de nuevas técnicas que permitan obtener un subconjunto óptimo de atributos.

Por otro lado, se está desarrollando una librería con el objetivo de facilitar el manejo de bases de datos de progenie en entornos de Big Data. El objetivo de esta librería es brindar una API simple para que cualquier investigador con pocos conocimientos de programación pueda utilizarla de manera simple.

En cuanto a las investigaciones relacionadas con el análisis de flujos de datos se centran en la construcción de modelos que faciliten la interpretación de los patrones obtenidos y la posterior extracción del conocimiento. En particular el énfasis está puesto en la resolución de dos problemas de sumo interés en distintas áreas: el mantenimiento de máquinas industriales basado en su condición de funcionamiento y el análisis de trayectorias GPS a fin de identificar las características del flujo vehicular en un período de tiempo dado.

Palabras clave: Big Data, Minería de Datos, Minería de procesos, Análisis de flujos de datos, Reducción de características.

1. INTRODUCCION

El Instituto de Investigación en Informática LIDI tiene una larga trayectoria en el estudio, investigación y desarrollo de Sistemas Inteligentes basados en distintos tipos de estrategias adaptativas. Los resultados obtenidos han sido aplicados en la solución de problemas de distintas áreas. A continuación, se detallan las

investigaciones realizadas durante el último año.

1.1. BIG DATA

Selección de características

En el área de la minería de datos y su aplicación con técnicas de machine learning, los algoritmos de selección de características juegan un papel muy importante. El objetivo de estos algoritmos es el de reducir las entradas a un tamaño apropiado para su procesamiento y análisis. La selección de características en una base de datos implica la elección de ciertos atributos, tal que, con ese subconjunto de atributos, las “propiedades naturales” de los datos pertenecientes a un dataset no sean alteradas.

Actualmente, en el III LIDI se están realizando tareas de investigación que incluyen el desarrollo de algoritmos de selección de características que puedan ser utilizados en bases de datos con información génica. Un objetivo de la medicina genómica es identificar un grupo de genes, cuyo patrón de expresión se encuentre asociado a un fenotipo en particular: concepto conocido como *gene signature* (biomarcador diagnóstico, pronóstico o predictivo de una patología en estudio [1]).

Cuando el volumen de información a procesar crece, la ejecución de los algoritmos de selección de atributos convencionales incrementa notablemente su tiempo de procesamiento. En la actualidad se cuenta con herramientas que, al distribuir el cómputo entre diferentes nodos que conforman un cluster de computadoras, hacen posible el procesamiento de grandes volúmenes de datos de una manera eficiente. En este aspecto Apache Spark es uno de los frameworks más utilizados. En particular, su librería Spark ML contiene la implementación de muchos algoritmos de machine learning.

En [2] se llevó a cabo una comparación entre cuatro algoritmos de clasificación implementados en MLlib: Random Forest, Support Vector Machine, Naïve Bayes y MultiLayer Perceptrón. Se analizaron cuatro métricas de poder pronóstico de los modelos junto con los tiempos de ejecución requeridos. Los experimentos están enfocados a comparar las métri-

cas de los cuatro algoritmos estudiados en función del número de atributos seleccionado de una base de datos.

Por otro lado, en un trabajo conjunto con el Centro de Altos Estudios en Tecnología Informática (CAETI) de la Facultad de Tecnología Informática (UAI) y el Centro de Investigaciones Inmunológicas Básicas y Aplicadas (CINIBA) de la Facultad de Ciencias Médicas (UNLP), se desarrollaron plataformas que permiten la consulta de bases de datos con información génica, en particular con micro-ARNs, moléculas de gran relevancia en estudios multi-ómicos que tienen por objetivo descifrar los mecanismos de desregulación de expresión génica en enfermedades complejas, multigénicas y multifactoriales, como es el caso de la mayoría de los tipos de cáncer.

En [3] se presenta Modulector, una plataforma que integra la información de las bases de datos públicas más relevantes de microARNs y las disponibiliza mediante una plataforma web simplificando el acceso a información actualizada. Esta plataforma puede ser accedida vía web a través de Multiomix [4], desde otras herramientas o plataformas que la integren, o incluso realizar un despliegue en un entorno privado, mediante el uso de las distintas APIs que provee.

Procesamiento de progenie en Big Data

Cuando en una base de datos, la información tiene naturaleza de un árbol n-ario, su análisis se dificulta, ya que para cualquier consulta deben realizarse varias operaciones *Join*. En particular, una operación por cada nivel del árbol que se desea explorar. El uso de múltiples *joins* incrementa la complejidad de la consulta a realizar, dificultando el análisis a quienes no tienen experiencia en programación.

En relación a esta línea, uno de los desarrollos que se están llevando a cabo es la implementación de una librería que permita el fácil tratamiento de los datos cuando estos están organizados y relacionados como un árbol, como por ejemplo, los datos de progenie.

En [6] presentamos TreeSpark, una herramienta basada en Spark, que permite realizar un

análisis de progenie. TreeSpark posee una API que permite un acceso simple a la información de los individuos y a sus relaciones tanto de progenie como de progenitores.

1.2. ANALISIS DE FLUJOS DE DATOS

Identificación de patrones de velocidad

El volumen de tráfico vehicular de las grandes ciudades se ha incrementado en los últimos años originando problemas de movilidad; por ello el análisis de los datos del flujo vehicular toma importancia para los investigadores. Los Sistemas Inteligentes de transporte realizan el monitoreo y control vehicular recolectando trayectorias GPS, información que brinda en tiempo real la ubicación geográfica de los vehículos. Por medio de técnicas de agrupamiento es posible identificar patrones sobre el flujo vehicular.

En este sentido se ha desarrollado en el III-LIDI una metodología capaz de identificar, de manera dinámica, las características del flujo vehicular en un período de tiempo. Para ello utiliza una adaptación original de una técnica de agrupamiento dinámico y un mapa interactivo para analizar el flujo del tráfico en sectores específicos facilitando la identificación de zonas de posibles atascos.

Los resultados obtenidos sobre un conjunto de datos de la ciudad de Guayaquil-Ecuador son satisfactorios y representan claramente la velocidad de desplazamiento de los vehículos identificando de manera automática los rangos más representativos para cada instante de tiempo. Como resultado del agrupamiento se obtuvieron diferentes grupos que dinámicamente cambian de una evolución a otra, identificando velocidades comunes en diferentes instantes de tiempo lo que permite tomar decisiones con respecto al tráfico de la ciudad. Para más detalle consultar [5] [7].

Deriva de concepto (concept drift)

El mantenimiento preventivo industrial se ha convertido en un tema importante y ha captado la atención de los investigadores en los últimos años. Las señales recogidas, como las

vibraciones, la corriente o las emisiones acústicas, son series temporales que suelen estar asociadas a una gran cantidad de flujos de datos. El análisis de estos flujos de datos se enfrenta a varios retos debido a que son potencialmente ilimitados, generados a ritmos muy rápidos, con limitaciones de tiempo de procesamiento y memoria y, además, pueden evolucionar con el tiempo.

La extracción de conocimiento a partir de estas grandes cantidades de datos asume que los datos son estáticos y generados por una distribución de probabilidad desconocida pero estacionaria. Sin embargo, en las aplicaciones de la vida real, la distribución de los flujos de datos cambia con el tiempo, ya que los entornos subyacentes son naturalmente dinámicos y se modifican constantemente. Este fenómeno se conoce como deriva de conceptos (concept drift). Los modelos que describen estos fenómenos deben actualizarse constantemente para reflejar el concepto actual. En el campo del mantenimiento preventivo, la deriva de concepto puede asociarse con los cambios en el estado de la máquina o en las condiciones de trabajo.

Recientemente, en el marco de una tesis de Doctorado en Cs. Informáticas de la UNLP se ha propuesto una metodología para detectar la evolución del concepto mediante el uso de medidas de divergencia para cuantificar las desviaciones. Fue utilizada en la clasificación de la severidad de los fallos de una caja de cambios helicoidal sometida a diversas cargas y velocidades. Los resultados muestran que el algoritmo propuesto detecta adecuadamente la evolución de concepto, que es un indicador de degradación de la máquina. La divergencia Kullback-Leibler ha resultado ser la métrica correcta para determinar la discrepancia entre las diferentes distribuciones de probabilidad empíricas dadas por las condiciones de fallo de la máquina.

Por otro lado, la selección de características en los flujos de datos sigue siendo un tema de investigación abierto. No hay muchos trabajos en torno al análisis de cómo evoluciona la importancia de las características con respecto al tiempo, por lo que se continuará trabajando

en esta importante parte de la metodología propuesta. Para más detalle consultar [8].

1.3. TINYML

TinyML es un campo de estudio que abarca al aprendizaje Automático y a los Sistemas Embebidos en la exploración de alternativas para adaptar modelos tradicionales para ejecutarlos en dispositivos con restricciones de memoria. En esta línea se inició un proyecto dedicado a la investigación de técnicas de Aprendizaje Automático aplicadas a microcontroladores (MCU) [10].

En el marco de este proyecto se construyeron modelos de redes neuronales convolucionales para analizar el funcionamiento de los frameworks Tensorflow Lite, Microtensor y Eloquent TinyML en dispositivos con diferentes características de hardware. Como resultado de este análisis se encontró que varios de estos frameworks requieren arquitecturas ARM, datos de 32 bits o soporte para instrucciones DSP o SIMD, dejando de lado a una cantidad importante de MCUs. Además, según estimaciones realizadas, se encontró que pequeños modelos requieren una cantidad de memoria mayor a la necesaria. En consecuencia, se decidió iniciar el desarrollo EmbedIA, un framework de código abierto compatible con C/C++ para realizar inferencia sobre modelos de redes neuronales convolucionales. Además de las funciones de inferencia se implementaron varias optimizaciones como aritmética de punto fijo para 32, 16 y 8 bits y estrategias de asignación de memoria para evitar la fragmentación. Se midió el desempeño de EmbedIA con los frameworks investigados en cinco microcontroladores diferentes [11][12]. En particular la implementación de punto fijo de 16 bits alcanzó, en promedio, una mejora de 5 a 10 veces en el tiempo de inferencia, de unas 3 veces en los requerimientos de memoria de datos y de 3 a 7 veces en los requerimientos de memoria de programa.

Actualmente, se está investigando la adaptación de otras técnicas de aprendizaje automático y posibles optimizaciones en la

implementación de redes neuronales convolucionales para reducir el impacto de la memoria requerida en microcontroladores.

2. TEMAS DE INVESTIGACIÓN Y DESARROLLO

- Medición de performance de algoritmos de machine learning en entornos distribuidos.
- Desarrollo de una plataforma pública de acceso web para ejecutar análisis de correlación entre grandes bases de datos de genes y moduladores de expresión.
- Desarrollo de una herramienta para el análisis de progenie, basada en Spark.
- Estudio de métricas para comparar distribuciones de probabilidad.
- Medidas de divergencia como herramienta para detectar la deriva de concepto.
- Preprocesamiento de trayectorias vehiculares. Técnicas de segmentación.
- Agrupamiento dinámico de flujos de datos.
- Estudio de técnicas de compresión de modelos para microcontroladores.
- Análisis de bibliotecas y frameworks de aprendizaje automático para microcontroladores.
- Desarrollo de un Framework que transforma modelos desarrollados en Tensorflow/Keras a código C/C++ para ejecutarlos en microcontroladores.

3. RESULTADOS OBTENIDOS

- Implementación de una plataforma web que integra información de bases de datos genómicas, con bases de datos de micro-ARNs, para simplificar el acceso a las distintas dimensiones de información.
- Implementación de una plataforma interactiva en la nube que permite identificar eventos (epi)genéticos asociados a la modulación transcripcional de genes relacionados con el cáncer a través del análisis de datos multiómicos.

- Implementación de una librería que facilita el desarrollo de pequeñas aplicaciones en Spark, para el tratamiento de bases de datos con información de proveniencia.
- Diseño e implementación de un nuevo método de detección de rangos de velocidad agrupamiento de trayectorias GPS aplicable a la predicción de congestiones vehiculares.
- Desarrollo de una metodología para detectar la evolución del concepto mediante el uso de medidas de divergencia para cuantificar las desviaciones.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo de la línea de I/D aquí presentada está formado por: 2 profesores doctores con dedicación exclusiva, 3 tesistas de Doctorado en Cs. Informáticas (1 con beca de postgrado de la UNLP), 1 tesista de grado y 2 profesores extranjeros.

Dentro de los temas involucrados en esta línea de investigación, en los últimos 3 años se han finalizado 4 tesis de doctorado, 3 tesis de especialista, 5 tesinas de grado de Licenciatura y 3 prácticas profesionales supervisadas.

Actualmente se están desarrollando 3 tesis de doctorado, 3 tesis de maestría, 4 tesis de especialista y 4 tesinas de grado de Licenciatura. También participan en el desarrollo de las tareas becarios y pasantes del III-LIDI.

5. REFERENCIAS

- [1] Abba, M. C.; Lacunza, E.; Butti, M. Breast cancer biomarker discovery in the functional genomic age: a systematic review of 42 gene expression signatures. *Biomarker Insights*; 5:1-16. 2010.
- [2] Camele, G.; Hasperué, W.; Ronchetti, F.; Quiroga, F.; A comparative study of the performance of four classification algorithms from the Apache Spark ML library. CACIC. Salta. 2021.
- [3] Marraco, A. D., Camele, G., Hasperué, W., Menazzi, S., Abba, M., & Butti, M. (2021). Modulector: una plataforma como servicio para el acceso a bases de datos de micro ARNs. *Innovación y Desarrollo Tecnológico y Social*, 3(1), 89-114.
- [4] Camele, G., Menazzi, S., Chanfreau, H., Marraco, A., Hasperué, W., Butti, M. D., & Abba, M. C. (2022). Multiomix: a cloud-based platform to infer cancer genomic and epigenomic events associated with gene expression modulation. *Bioinformatics*, 38(3), 866-868.
- [5] Reyes G., Lanzarini L., Fernández Bariviera A., Hasperué W. A proposal for a pivot-based vehicle trajectory clustering method. *Transportation Research Record, Journal of the Transportation Research Board*. ISSN: 0361-1981.
- [6] López, P. D.; Hasperué, W.; Quiroga, F.; Ronchetti, F. TreeSpark: A Distributed Tool for Progeny Analysis based on Spark. CACIC. Salta. 2021.
- [7] Reyes G., Lanzarini L., Estrebou C. (2021). Vehicular Flow Analysis Using Clusters. XVIII Workshop Base de Datos y Minería de Datos (WBDMD) en CACIC. Univ. Nacional de Salta. Octubre 2021.
- [8] Peña M., Lanzarini L., Cerrada M., Sánchez RV. (2021) Data-Driven Gearbox Fault Severity Diagnosis Based on Concept Drift. IEEE Fifth Ecuador Technical Chapters Meeting (ETCM), 1-6
- [10] Estrebou C., Feming M., Saavedra M. D., Adra F. MbedML: A Machine Learning Project for Embedded Systems. IX Jornadas de Cloud Computing, Big Data & Emerging Topics. ISBN: 978-950-34-2016-4 Pp. 25-28. Facultad de Informática. UNLP. Junio 2021.
- [11] Estrebou C., Feming M., Saavedra M. D., Adra F. A Neural Network Framework for Small Microcontrollers. XXVII Congreso Argentino de Ciencias de la Computación. ISBN: 978 -987-633-574-4. Pp. 51-60. Univ. Nacional de Salta. Octubre 2021.
- [12] Estrebou C., Feming M., Saavedra M. D., Adra F., De Giusti, A. E. Lightweight Convolutional Neural Networks

Framework for Really Small TinyML Devices. Second International Conference on Smart Technologies, Systems and Applications. SmartTech-IC 2021. Quito, Ecuador. Diciembre 2021.

-
-  Waldo Hasperué: 0000-0002-9950-1563
 -  César Estrebou: 0000-0001-5926-8827
 -  Genaro Camele: 0000-0001-6979-9103
 -  Mario Peña: 0000-0002-3986-7707
 -  Gary Reyes: 0000-0002-3711-1906
 -  Laura Lanzarini: 0000-0001-7027-7564
 -  Aurelio Fernandez Bariviera:
0000-0003-1014-1010
 -  Mariela Cerrada: 0000-0003-4379-8836

Aplicaciones de Bases de Datos Espaciales y Espacio Temporales

Edilma Olinda Gagliardi, Maria Gisela Dorzán, Maria Teresa Taranilla,

Pablo Rafael Palmero y Carlos Andrés Casanova

Departamento de Informática

Facultad de Ciencias Físico Matemáticas y Naturales

Universidad Nacional de San Luis, Argentina

{oli, mgdorzan, tarani, prpalmero, cacasanova}@unsl.edu.ar

RESUMEN

En la línea de investigación *Bases de Datos Espaciales y Espacio Temporales* perteneciente al proyecto *Tecnologías Avanzadas de Bases de Datos* desarrollado en la Universidad Nacional de San Luis, se exploran dominios de aplicación de bases de datos espaciales y espacio temporales, utilizando técnicas, métodos y herramientas para investigación de base, así como el desarrollo de aplicaciones para contribuir en la resolución de problemas provenientes de diferentes ámbitos de aplicación.

Palabras clave: Bases de Datos, Bases de Datos Espaciales y Espacio Temporales, Geometría Computacional, AgroTIC, Gestión Integral Incendios Forestales.

CONTEXTO

En el marco del Proyecto *Tecnologías Avanzadas de Bases de Datos* (PROICO N°03-2218) incluido en el Programa de Incentivos (código 22/F814) y en el contexto del *Laboratorio de Investigación y Desarrollo en Bases de Datos (LABDA)* de la Universidad Nacional de San Luis, se enmarca el trabajo de la línea de investigación *Bases de Datos Espaciales y Espacio Temporales*.

El proyecto está conformado por tres líneas de investigación orientadas al estudio de bases de

datos no convencionales y al desarrollo de nuevos modelos para administrar y recuperar información almacenada en repositorios de datos no estructurados.

En particular, los modelos de bases de datos espaciales y espacio temporales se utilizan en contextos donde es necesario almacenar y consultar información actual e histórica de posiciones referenciadas espacialmente y, además, poder observar cambios de forma de los objetos de estudio en diferentes escenarios a lo largo del tiempo.

La línea de investigación *Bases de Datos Espaciales y Espacio Temporales* involucra el diseño y desarrollo de herramientas para administrar y extraer información de sistemas de bases de datos. En este sentido, se propone el estudio de herramientas teóricas y diseño de aplicaciones que permitan modelar y administrar datos, organizarlos y realizar operaciones de interés sobre ellos, como así también obtener información para análisis complejos, descubrimiento de conocimiento de apoyo en la toma de decisiones en un sistema.

El trabajo de investigación se desarrolla en estrecha colaboración con investigadores afines de proyectos de la Universidad Nacional de San Luis, y a través de convenios entre organizaciones nacionales y provinciales con presencia en San Luis.

1. INTRODUCCIÓN

Las Tecnologías de la Información y la Comunicación (TIC) y su inclusión en la vida cotidiana han impactado en la mayoría de las actividades humanas, promoviendo cambios sociales, culturales y económicos que repercuten en procesos clave que involucran toma de decisiones en el presente y para proyectar el futuro. Las TIC aportan beneficios en diferentes situaciones del mundo real y proporcionan mejoras en tópicos que refieren a prevención, gestión, calidad, economía, entre otros. En múltiples contextos se perfilan experiencias que permiten innovar en nuevas formas de comunicación y en prácticas orientadas a optimizar tareas relacionadas con la gestión de diversas temáticas.

En este sentido, los integrantes de la línea integraron el proyecto Campo Conectado, resultado de un convenio entre el Instituto Nacional de Tecnología Agropecuaria (INTA-Estación Experimental Agropecuaria San Luis) y la Universidad Nacional de San Luis con el objetivo promover un ámbito de investigación, intercambio y desarrollo y de estudiar la potencialidad y alcance real de las TIC en el ámbito de la producción agropecuaria [1] [2].

Por otra parte, desde el mes de octubre de 2021 el grupo de trabajo integra el Programa “Prevención y Gestión Integral de Incendios Forestales” en el marco del *Programa de Investigación Institucional 2021-2022* de la Universidad Nacional de San Luis en el área estratégica Medio Ambiente. Estos programas creados según Resolución C.S 33/20-UNSL definen entre sus propósitos abordar problemas y temas de investigación institucionales en un trabajo articulado e interdisciplinario. Además de fomentar el desarrollo de la inter e intra institucionalidad, cooperación e intercambios en la función I+D+i.

Los incendios forestales frecuentes tienen un tremendo impacto negativo sobre los ecosistemas nativos al interrumpir los ciclos naturales de las especies vegetales y animales, provocando la desaparición de las especies nativas, favoreciendo las invasiones biológicas y degradando el suelo en sus propiedades

físico-químicas y biológicas. Es importante considerar el impacto económico que un incendio forestal tiene sobre los recursos intangibles como el valor del bosque para el ocio, valor paisajístico o la fijación de carbono. Así como también, el deterioro veloz de varios ecosistemas únicos, con menor resiliencia para la recuperación del paisaje tras el incendio.

El programa “Prevención y Gestión Integral de Incendios Forestales” propone un plan de estudio interdisciplinario para la prevención y gestión integral de incendios forestales, a través de la interacción entre la investigación básica y aplicada, con el propósito de diseñar instrumentos y/o prototipo de soporte a diversos eventos y sistemas que contribuyan a la toma de decisiones y análisis de impacto [10].

En algunos de los problemas que se plantean en los dominios de aplicación previamente citados, surge la necesidad de representar y manipular tipos de datos tradicionales y tipos de datos complejos. Existen requerimientos de almacenamiento y consultas de información actual e histórica y acerca de objetos de estudio referenciados espacialmente y su evolución en diferentes escenarios a lo largo del tiempo.

En este sentido, las Bases de Datos Espaciales permiten representar y almacenar objetos junto con sus características espaciales, así como expresar consultas que involucran sus propiedades espaciales. Las Bases de Datos Temporales involucran el aspecto tiempo para organizar su información y almacenan datos históricos junto con los datos actuales. Las Bases de Datos Espacio Temporales concentran características de ambas y permiten observar la evolución de objetos espaciales en el tiempo [4] [8] [11]. En esta línea de trabajo se exploran dominios de aplicación para bases de datos espaciales y espacio temporales, utilizando diferentes técnicas y herramientas en la resolución de problemas. Se plantea el diseño y aplicación de índices espacio temporales, aplicables en diversos escenarios y se propone la utilización de técnicas de

Geometría Computacional en la solución de los problemas involucrados [3].

2. LÍNEA DE INVESTIGACIÓN

En la línea se investigan entornos de aplicación de bases de datos espaciales y espacio temporales, con uso de técnicas y herramientas de apoyo en la resolución de problemas.

Entre los tópicos de estudio de la línea de investigación se destaca:

- Diseño y aplicación de índices espaciales y espacio temporales, en diversos escenarios de movimiento.
- Desarrollo de herramientas y aplicaciones vinculadas con bases de datos espaciales y espacio temporales.
- Aplicación de la Geometría Computacional y su marco disciplinar en aspectos particulares de los problemas estudiados.

Como objetivos específicos se propone:

- Estudio de la indexación espacial y espacio temporal sobre objetos en movimiento para diversos escenarios. Desarrollo de las estructuras de almacenamiento, algoritmos de consulta y evaluación experimental.
- Desarrollo de aplicaciones que utilizan bases de datos espaciales y espacio temporales, con apoyo de métodos provistos por la Geometría Computacional en la resolución de situaciones problemáticas.
- Diseño e implementación de herramientas para la visualización de información de interés vinculada a las bases de datos mencionadas.

3. RESULTADOS OBTENIDOS

En el marco del proyecto Campo Conectado, se han propuesto herramientas tecnológicas de apoyo para la gestión de la producción agropecuaria.

En la producción ganadera, algunos de los requerimientos relevados refieren al seguimiento de indicadores productivos y ambientales, caracterización y seguimiento de individuos de un rodeo, trazabilidad animal, control de alimentación, detección de la actividad de celo, monitoreo de la condición corporal, entre otros, como herramientas en la ganadería de precisión. En el marco de estos requerimientos se propuso una plataforma soporte de eventos y sistemas de información orientada al seguimiento espacio temporal de rodeos y la gestión espacial y tradicional de los elementos que integran un establecimiento agropecuario [5].

En la plataforma integral se incluye una herramienta para el seguimiento espacial y temporal de individuos de rodeos ganaderos y una herramienta de apoyo para la gestión de rodeos de cría.

La herramienta para el seguimiento espacio temporal de rodeos manipula las componentes espaciales y tradicionales de los datos, implementa consultas espaciales y espacio temporales sobre los objetos de tratamiento, proporciona comunicación con las bases de datos y visualización del geo seguimiento del rodeo por medio de tecnología móvil [6] [9].

La herramienta para la gestión de rodeos de cría incluye funcionalidades para el análisis de datos relevados, seguimiento de indicadores como la condición corporal de los individuos del rodeo y la obtención de conocimiento de apoyo con su correspondiente visualización disponible con tecnologías móviles [7].

En el marco Programa “*Prevención y Gestión Integral de Incendios Forestales*”, se está trabajando en estrecha colaboración con los integrantes de seis proyectos pertenecientes a tres Unidades Académicas (FCFMyN, FQBF y FICA) de la UNSL, que conforman este programa.

Desde la línea de investigación se participa con el aporte en el diseño y asistencia técnica en el desarrollo de bases de datos espaciales, procesamiento y optimización de consultas a bases de datos existentes asociadas a la problemática y con la contribución en la

definición de indicadores temáticos a los efectos de presentar información de mayor relevancia para el seguimiento y análisis de las transformaciones ambientales, que permitan dar apoyo en la toma de decisiones referidas a prevención, gestión e impacto de incendios forestales, así como la recuperación ambiental.

Como ejes de trabajo futuros, se propone continuar con actividades que involucran la pertinencia y transferencia de proyectos I+D+i en las áreas de intervención seleccionadas, maximizando la sinergia entre grupos de trabajo, afianzando el ámbito interdisciplinario de investigación y con el fortalecimiento de competencias.

4. FORMACIÓN DE RECURSOS HUMANOS

Los integrantes de la línea de investigación consolidan su formación con actividades de cooperación mutua e intercambio recíproco de información científica, tecnología y desarrollo de nuevos conocimientos con investigadores locales y de otras universidades.

Las actividades de formación de recursos humanos, se reflejan en la dirección de tesis de maestría y trabajos finales de Licenciatura en Ciencias de la Computación, dirección de becas y dictado de cursos de postgrado. Además, se realizan actividades de formación académica, con la realización de cursos de posgrado y de especialización, participación en actividades de divulgación científica, conferencias, y otras actividades académico-científicas vinculantes.

5. BIBLIOGRAFÍA

- [1] *Agrotic: Revolución Digital, Desarrollo y Producción*.
<https://inta.gob.ar/noticias/agrotic-revolucion-digital-desarrollo-y-produccion>
(2016)
- [2] *Campo Conectado*, proyecto interinstitucional
<http://inta.gob.ar/noticias/campo-conectado-un-nuevo-proyecto-interinstitucional> (2017)
- [3] de Berg, M., Cheong O., van Kreveld, M., Overmars, M. : *Computational Geometry: Algorithms and Applications.*, Springer-Verlag, Heidelberg (2008)
- [4] Elmasri, R. Navathe, S.B.: *Fundamentals of Database Systems*. 7^{ma} Edition. Pearson (2015)
- [5] Gagliardi, E., Dorzán, M. G., Taranilla, M. T., Palmero, P., Casanova, C.: *Propuesta de plataforma para la integración de TIC orientadas al Agro* Congreso de AgroInformática CAI 2017, 46JAIIO (2017)
- [6] Gagliardi, E.; Dorzán, M.G.; Taranilla, M.T.; Palmero, P.; Casanova, C.: *GeoSeguimiento de Rodeos, hacia una plataforma integral para el Agro*. Congreso de AgroInformática CAI 2018, 47JAIIO (2018).
- [7] Gagliardi, E.; Dorzán, M.G.; Taranilla, M.T.; Palmero, P.; Casanova, C.: *Diseño de una herramienta de apoyo en la gestión de rodeos de cría*. Congreso de AgroInformática CAI 2018, 47JAIIO (2018).
- [8] Manolopoulos, Y., Papadopoulos, A., Vassilakopoulos, M.: *Spatial Databases: Technologies, Techniques and Trends*. Idea Group (2005)
- [9] Palmero, P.; Gagliardi, E.; Dorzán, M.G.: *Seguimiento de rodeos en establecimientos agropecuarios*. CACIC: XXVI Congreso Argentino de Ciencias de la Computación (2020)
- [10] Programa “*Prevención y Gestión Integral de Incendios Forestales*”, Programa de Investigación Institucional 2021-2022 en el área estratégica Medio Ambiente. Resolución N° CS 237-21 - UNSL
http://digesto.unsl.edu.ar/docs/202110/20211006125521_1112.pdf
- [11] Shekhar, S.; Chawla, S.: *Spatial databases: a tour*. Prentice Hall (2003)

Indexación y Búsquedas sobre Datos no Estructurados

Norma Herrera, Darío Ruano, Paola Azar

Departamento de Informática
Universidad Nacional de San Luis, Argentina
{nherrera, dmruano, epazar}@unsl.edu.ar

Anabella De Battista, Andrés Pascal

Departamento Ingeniería en Sistemas de Información
FRCU, Universidad Tecnológica Nacional, Entre Ríos, Argentina
{anadebattista, andrespascal22}@gmail.com

Resumen

Debido al crecimiento exponencial de las fuentes de información disponibles, en la actualidad resulta necesario contar con técnicas y herramientas diferentes a las tradicionales para abordar el manejo de la información. En la actualidad existen bases de datos no estructurados que no son asimilables con los modelos clásicos de bases de datos, como el modelo relacional. De aquí surge la necesidad de contar con nuevos modelos de datos y nuevas herramientas para el manejo de los mismos.

En este proyecto nuestro interés está centrado al estudio de modelos para bases de datos no estructurados y de técnicas de indexación asociados a los mismo.

1 Contexto

El presente trabajo se desarrolla en el ámbito de la línea Técnicas de Indexación para Datos no Estructurados del Proyecto Tecnologías Avanzadas de Bases de Datos (22/F814), cuyo objetivo es realizar investigación básica sobre manejo y recuperación eficiente de información no tradicional.

2 Introducción

Los avances en las tecnologías de la información y la comunicación han permitido que las bases de datos crezcan de manera exponencial en tamaño y se diversifiquen en contenido. Gran parte de la información disponible en la actualidad involucra el uso de datos no estructurados. Por no estructurados entendemos un conjunto de datos que no admiten la representación clásica de bases de datos relacionales, donde la información se organiza como relaciones (tablas) cada uno compuesta por nuplas (filas de la tabla), y donde cada nupla tiene un valor para cada atributo de la relación (columnas de la tabla). Como ejemplos de bases de datos no estructurados podemos mencionar bases de datos de imágenes, de sonidos, de textos, de huellas digitales, entre otras.

Mientras que en base de datos relacionales la búsqueda más común es la búsqueda exacta, en bases de datos no estructurados la búsqueda exacta carece de interés. Surgen otras necesidades de almacenamiento y consulta que no están consideradas en los modelos clásicos de bases de datos.

Es en este contexto donde surgen nuevos modelos de bases de datos capaces de cubrir las necesidades de las aplicaciones que manejan datos no estructurados. Entre esos modelos encontramos el modelo de *bases de datos de texto* y el modelo *métrico-temporal*, sobre los que actualmente nos encontramos trabajando. Damos a continuación una breve descripción de cada uno de ellos.

Bases de Datos de Texto

Una base de datos de texto es una gran colección de texto sobre la que se requiere resolver consultas de manera eficiente. Sin pérdida de generalidad asumiremos que la base de datos es un gran texto T , posiblemente dividido en varios archivos. Este texto T puede representar no sólo lenguaje natural, sino también música, códigos de programas, secuencias de ADN, secuencias de proteínas, etc.

Una de las búsquedas más comunes en bases de datos de texto es la *búsqueda de un patrón*: el usuario ingresa un string P (*patrón de búsqueda*) y el sistema retorna todas las posiciones del texto T donde P ocurre. Resolver la búsqueda de un patrón de manera eficiente en grandes colecciones de texto requerirá la construcción de un índice.

Mientras que en bases de datos tradicionales los índices ocupan menos espacio que el conjunto de datos indexados, en bases de datos de texto el índice ocupa más espacio que el texto en sí mismo, pudiendo necesitar de 4 a 20 veces el tamaño del mismo [7][10]. Esto implica que un índice construido sobre una base de datos de texto residirá en memoria secundaria y en consecuencia la cantidad de accesos a disco realizados durante el proceso de búsqueda será un factor crítico en la performance del mismo [14]. Por esta razón, el diseño de técnicas de paginado eficientes es de vital importancia para índices en memoria secundaria.

Bases de Datos Métrico-Temporales

En el caso del modelo **métrico-temporal** [11, 2] fue pensado para manejar objetos no estructurados con tiempos de vigencia asociados, y permitir realizar consultas por similitud y por tiempo en forma simultánea. Formalmente un *espacio métrico-temporal* es un par (U, d) , donde $U = O \times N \times N$, y la función d es de la forma $d : O \times O \rightarrow R^+$. Cada elemento $u \in U$ es una tripla (obj, t_i, t_f) , donde obj es un objeto (por ejemplo, una imagen, sonido, cadena, etc) y $[t_i, t_f]$ es el intervalo de vigencia de obj . La función de distancia d , que mide la similitud entre dos objetos, cumple con las propiedades de una métrica (positividad, simetría y desigualdad triangular).

Una *consulta métrico-temporal* se define como: $(q, r, t_{iq}, t_{fq})_d = \{o / (o, t_{io}, t_{fo}) \in X \wedge d(q, o) \leq r \wedge (t_{io} \leq t_{fq}) \wedge (t_{iq} \leq t_{fo})\}$ Esta consulta implica buscar todos los objetos o de la base de datos que estén a una distancia a lo más r de q , y cuyo tiempo asociado t se solapa con el tiempo de la consulta.

Varios índices métrico-temporales se han propuesto en este ámbito, todos estos índices fueron desarrollados para ser eficientes en memoria principal.

3 Líneas de Investigación

Las líneas I+D de este proyecto se concentran en el diseño de índices eficientes para bases de datos no estructurados. Describimos a continuación las principales líneas en las que nos encontramos trabajando actualmente.

Tries de Sufijos en Memoria Principal

En este ámbito, usando como base el trie de sufijos [7], el objetivo final es el diseño de un índice en memoria secundaria, que además sea eficiente en espacio. El proceso de paginación de un índice consiste en dividir el mismo en partes, cada una de las cuales se

aloja en una página de disco. Luego el proceso de búsqueda consiste en ir cargando en memoria principal una parte, realizar la búsqueda en memoria principal sobre esa parte, para luego cargar la siguiente y proseguir la búsqueda. Cuando un índice se maneja en disco, el costo de búsqueda queda determinado por la cantidad de accesos a disco realizadas [15]. Aun así, es importante no descuidar las operaciones que se hacen en memoria principal a fin de lograr un funcionamiento eficiente del índice. Es por esta razón que es necesario evaluar el desempeño en memoria principal de la representación que utilizaremos para la versión en memoria secundaria.

En una primera etapa estudiamos el desempeño en memoria principal de distintas variantes de representación del trie de sufijos, que se adecuaran a un posterior proceso de paginado del índice. En [12] se propuso y evaluó una representación secuencial del trie de sufijos (que denotaremos con TS_s) que cumple con este requisito. En [13] se presentó una mejora en espacio del trie de sufijos, denotada con TS_{dac} , que consiste en el uso de códigos DAC (Directly Addressable Codes [3]) en algunas componentes de TS_s . En [5] se propuso una mejora en tiempo del mismo índice (que denotaremos con TS_{dacn}) que implica agregar una componente más a TS_s para mejorar las operaciones que se necesitan realizar durante la búsqueda de un patrón. Con esta última versión se lograron mejoras significativas en los tiempos de búsqueda usando muy poco espacio adicional.

Un punto importante a tener en cuenta aquí es el algoritmo de creación de este índice en memoria principal. Existen varias propuestas al respecto, nosotros hemos seleccionado el algoritmo propuesto en [9] que consiste en construir este índice a partir del arreglo de sufijos A y del arreglo LCP (longest common prefix array)[10]. Este algoritmo de creación utiliza como estrategia un diseño bottom-up, comenzando por las hojas para luego ir subi-

endo en la construcción del trie hasta llegar a la raíz, usando como estructura auxiliar una pila P . La idea general del algoritmo es insertar las hojas en el trie de izquierda a derecha, usando para ello la información brindada por el arreglo de sufijos A . El lugar de inserción de una hoja $A[i]$ se decide en función del valor $LCP[i]$. Cada vez que insertamos una nueva hoja $A[i]$, ésta se convertirá en el hijo de más a la derecha de un nodo v , sobre el camino de más a la derecha del trie. En la pila P se mantiene esta información, es decir cada elemento de P representa un nodo del camino de más a la derecha, el cual se reconstruye recorriendo los elementos de P desde la base hacia el tope. Por lo tanto el valor que se encuentra en la posición p de P es hijo del nodo que se encuentra en la posición $p - 1$ de la misma, encontrándose en el tope de P la última hoja insertada.

Tries de Sufijos en Memoria Secundaria

Uno de los objetivos del trabajo desarrollado hasta el momento es lograr una versión en memoria secundaria de TS_{dacn} .

Hemos diseñado una técnica de paginado que consiste en particionar el árbol en componentes conexas, denominadas *partes*, cada una de las cuales debe tener un tamaño que no supere la capacidad de una página de disco. El algoritmo procede en forma bottom-up tratando de condensar en una única parte un nodo con uno o más subárboles que dependen de él. En este proceso de particionado las decisiones se toman en base a la profundidad de cada nodo involucrado, donde la profundidad indica la cantidad de accesos a disco que deberá realizar el proceso de búsqueda para llegar desde esa parte a una hoja del árbol. Este algoritmo de paginado es una extensión a árboles r -arios del algoritmo presentado en [4] para un árbol binario. Dicho algoritmo es *min-max óptimo*, es decir que minimiza la cantidad máxima de accesos a páginas de disco necesarias para recorrer el árbol desde la raíz hasta

una hoja.

Existen dos posibilidades para la creación de TS_{dact} en memoria secundaria. El método ingenuo consistiría en crearlo con el algoritmo de creación para memoria principal explicado anteriormente, para después proceder a paginarlo con la técnica diseñada. Claramente esto provocará que la creación sea ineficiente, dado que estamos dejando al sistema de memoria virtual del sistema operativo el manejo de la memoria necesitada por el índice que estamos creando. Para evitar esto lo que proponemos es compatibilizar el algoritmo de creación del trie del sufijo en memoria principal con el algoritmo de paginado diseñado. Ya se ha diseñado en detalle este algoritmo y nos encontramos en la etapa de codificación del mismo.

Bases de Datos Métrico-Temporales

En este ámbito, varios índices fueron propuestos para la resolución eficiente de consultas métrico-temporales. Todos estos índices fueron desarrollados para ser competitivos en memoria principal.

El *Historical-FHQT*(H-FHQT) [6] es un índice métrico-temporal que utiliza tanto la componente métrica como la temporal para resolver eficientemente búsquedas métrico-temporales. Este índice consiste en una lista de instantes válidos donde cada uno contiene un FHQT [1] que indexa a todos los objetos vigentes en dicho instante. Los FHQT tienen distintas profundidades, es decir distintas cantidades de pivotes, en función de la cantidad de elementos que deban indexar.

Si bien en un H-FHQT la cantidad de pivotes en distintos instantes de tiempo varía, siempre se trabaja sobre el mismo conjunto base de pivotes; esto significa que si en el instante i necesito k_i pivotes y en el instante j necesito k_j pivotes con $k_i < k_j$, entonces los primeros k_i pivotes de ambos instantes son iguales.

En este ámbito nuestro objetivo es el diseño de una versión en memoria secundaria del H-FHQT. Hemos diseñado ya una primer

técnica de paginación de este índice y los resultados han sido alentadores (detalles de la misma pueden consultarse en [8]). Como próximo paso nos proponemos mejorar la técnica diseñada permitiendo que en las decisiones de paginado se utilice también los rótulos de la ramas (valores de distancia) de cada FHQT involucrado.

4 Resultados Esperados

Como resultados de los trabajos que estamos realizando, esperamos obtener índices eficientes en memoria secundaria para las bases de datos planteadas. Todo el trabajo realizado nos permite adquirir la experiencia suficiente como para abocarnos al diseño de otros índices que sean competitivos en este ámbito.

Cabe señalar, que en el caso de las bases de datos métrico-temporales, no existe un core de datos con los cuales realizar los experimentos. Por lo tanto, un resultado adicional será la generación de lotes de pruebas voluminosos que puedan servir como referencia en este ámbito de estudio.

5 Recursos Humanos

Dentro de esta línea se forman docentes y alumnos de la Universidad Nacional de San Luis y de la Universidad Tecnológica Nacional (FRCU). Actualmente hay en desarrollo 2 Tesis de Maestría y un Trabajo Final de la Licenciatura, todos de la Universidad Nacional de San Luis.

Referencias

- [1] R. Baeza-Yates, W. Cunto, U. Manber, and S. Wu. Proximity matching using fixed-queries trees. In *Proc. 5th Combinatorial Pattern Matching (CPM'94)*, LNCS 807, pages 198–212, 1994.

- [2] A. De Battista, A. Pascal, N. Herrera, and G. Gutierrez. Metric-temporal access methods. *Journal of Computer Science & Technology*, 10(2):54–60, 2010.
- [3] N. Brisaboa, S. Ladra, and G. Navarro. Directly addressable variable-length codes. In *Proc. 16th International Symposium on String Processing and Information Retrieval (SPIRE)*, LNCS 5721, pages 122–130. Springer, 2009.
- [4] D. Clark and I. Munro. Efficient suffix tree on secondary storage. In *Proc. 7th ACM-SIAM Symposium on Discrete Algorithms*, pages 383–391, 1996.
- [5] J. Cornejo, D. Ruano, and N. Herrera. Una mejora en tiempo del trie de sufijos. In *Congreso Argentino de Ciencias de la Computación*, Rio Cuarto, Córdoba, Argentina, 2019.
- [6] A. De Battista, A. Pascal, G. Gutierrez, and N. Herrera. Un nuevo índice métrico-temporal: el historical-fhqt. In *Actas del XIII Congreso Argentino de Ciencias de la Computación*, Corrientes, Argentina, 2007.
- [7] G. H. Gonnet, R. Baeza-Yates, and T. Snider. *New indices for text: PAT trees and PAT arrays*. Prentice Hall, New Jersey, 1992.
- [8] N. Herrera, D. Ruano, P. Azar, Daniel Welch, L. Speranza, M. de la Torre, N. A. De Battista, and Andrés Pascal. Técnicas de indexación para bases de datos avanzadas. In *Proc. XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021)*, pages 240–244, Chilecito, La Rioja, Argentina, 2021.
- [9] Juha Kärkkäinen and S. Srinivasa Rao. Full-text indexes in external memory. In Ulrich Meyer, Peter Sanders, and Jop F. Sibeyn, editors, *Algorithms for Memory Hierarchies*, volume 2625 of *Lecture Notes in Computer Science*, pages 149–170. Springer, 2002.
- [10] U. Manber and G. Myers. Suffix arrays: A new method for on-line string searches. *SIAM Journal of Computing*, 22(5):935–948, 1993.
- [11] A. Pascal, A. De Battista, G. Gutierrez, and N. Herrera. Métodos de acceso para bases de datos métrico - temporales. In *Actas del XV Congreso Argentino de Ciencias de la Computación*, pages 1061–1070, Jujuy, Argentina, 2009.
- [12] D. Ruano and N. Herrera. Representación secuencial de un trie de sufijos. In *XX Congreso Argentino de Ciencias de la Computación*, Buenos Aires, Argentina, 2014.
- [13] D. Ruano and N. Herrera. Indexando bases de datos textuales: Una representación compacta del trie de sufijos. In *Congreso Nacional de Ingeniería Informática / Sistemas de Información*, Buenos Aires, Argentina, 2015.
- [14] J. Vitter. External memory algorithms and data structures: Dealing with massive data. *ACM Computing Surveys*, 33(2):209–271, 2001.
- [15] Jeffrey Scott Vitter. *Algorithms and Data Structures for External Memory*. Publishers Inc, 2006.

MÉTODOS DE SERIES DE TIEMPO DE INTERVALO

Adriana Mallea¹, Cecilia Martínez²,
Andrea Salas¹

¹Departamento de Matemática, FFHA, Universidad Nacional de San Juan

²Departamento de Informática, FCFN, Universidad Nacional de San Juan

lamallea@ffha.unsj.edu.ar

RESUMEN

El objetivo principal de esta línea de investigación es mostrar el desarrollo de métodos de predicción para series temporales de intervalo (ITS) y su aplicación en ejemplos reales para demostrar su efectividad.

Este plan surge de la confluencia de dos disciplinas: el análisis de datos simbólicos (área que se ocupa del análisis estadístico de datos de intervalo, en este caso) y la predicción de series temporales. De acuerdo a esto y para lograr el objetivo general, se proponen los siguientes objetivos particulares:

1. Estudio de métodos de predicción de series temporales simbólicas para datos de intervalo. Proponemos estudiar métodos propuestos en la literatura, siguiendo dos enfoques:

- Predecir la serie mediante métodos de predicción clásicos: consiste en tratar la serie temporal de intervalos como un par de series temporales clásicas que pueden ser la serie de los mínimos y la de los máximos o, alternativamente, la de los centros y la de los radios.
- Predecir la serie mediante métodos de predicción que consideren el intervalo como tal, todos ellos basados en la

aritmética de intervalos (Moore, 1966)

2. Estudio de modelos vectoriales autorregresivos de promedios móviles para series de intervalo.

Recientemente, en la literatura de pronóstico de ITS, se ha propuesto el modelo Vectorial Autorregresivo de Promedios Móviles (IVARMA) para capturar la dinámica de cross-dependencia presente en este tipo de series.

3. Aplicación de los métodos de predicción estudiados a series provenientes de diferentes ámbitos.

Se ilustrará la metodología estudiada en series clásicas provenientes de diversas áreas de la actividad humana, sobre todo en contextos econométricos. Para ello se procederá a la agregación temporal, en cada caso, de los datos unvariados a fin de lograr el intervalo de valores en instantes específicos de tiempo, se aplicarán los distintos métodos para pronóstico de ITS estudiados y se evaluará su performance con el error definido como la raíz cuadrada del RECM (error cuadrático medio escalado).

Palabras clave: Series, Tiempo, intervalo.

CONTEXTO

Esta línea de investigación se inserta en el proyecto *Análisis de Datos Simbólicos para Data Science*, cuyo tipo de actividad de I+D es investigación básica, que inició en el año 2020 y finaliza en diciembre de 2022, financia la UNSJ. Tiene como unidad ejecutora el Departamento de Matemática de la FFHA y sus integrantes desarrollan sus tareas de docencia e investigación en las áreas de matemática e informática.

1. INTRODUCCIÓN

Los datos simbólicos son un paradigma de representación de la información que surge a finales de los ochenta (Diday, 1987) bajo la premisa de que las variables clásicas, i.e., aquellas que a cada individuo le asignan un único valor, no son capaces de representar con fidelidad algunas situaciones. Los datos simbólicos, a diferencia de los clásicos, permiten representar conceptos de una manera sintética y descriptiva.

La característica fundamental de los datos simbólicos es que permiten la descripción de elementos o fenómenos donde exista una variabilidad interna.

Los conceptos implican variabilidad ya que las distintas realizaciones de ese concepto pueden ser algo diferentes entre sí. La variabilidad surge de manera natural al agregar observaciones. Por agregación se entiende la recopilación de observaciones que satisfacen un requisito que les permite ser agrupadas.

La agregación puede ser:

- Contemporánea, si se recopilan observaciones recogidas en un mismo instante temporal o cuando el instante temporal no es relevante.
- Temporal, si el criterio de agregación es el tiempo y se

recopilan observaciones ocurridas a lo largo de una unidad de tiempo, e.g. un día.

Al tener una estructura distinta que la de los datos clásicos, las técnicas de análisis del paradigma clásico no son válidas para analizar los datos simbólicos. Por ello, es necesario desarrollar un nuevo catálogo de métodos que sean capaces de extraer el conocimiento de este nuevo tipo de datos. Éste es el propósito del análisis de datos simbólicos. Este tipo de análisis es un área novedosa dentro de la estadística y de la minería de datos. Su objetivo es el de extraer información sobre datos expresados mediante variables simbólicas, como por ejemplo variables de intervalo. Estas variables son más complejas que las variables clásicas que describen los elementos mediante un único valor, pero describen de manera más fiel la realidad al poder representar variabilidad.

A lo largo de las últimas tres décadas las técnicas del Análisis de Datos Simbólicos (ADS) se han incrementado notablemente. En esta línea de investigación, en particular, interesa mostrar el avance de contribuciones en el análisis de series temporales de intervalo.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

La línea de investigación se enmarca dentro de Data Science y Data Mining aplicados a los datos estándar.

Los primeros trabajos para el pronóstico de ITS que se proponen en la literatura son un modelo ARMA y un modelo híbrido que combina el modelo ARMA con una red neuronal artificial (RNA). Teles y Brito (2005) proponen una extensión de los modelos ARMA para trabajar con series temporales de intervalos. Ellos asumen que los procesos que generan los mínimos y los máximos

de los intervalos tienen los mismos parámetros pero diferentes medias, siendo la media de la ecuación de los máximos mayor que la media de la ecuación de los mínimos. En Maia et al. (2006a) y Maia, de Carvalho y Ludermir (2006b), los autores proponen una extensión de los modelos ARMA y de los modelos híbridos para predecir series temporales de intervalo. La extensión de los modelos ARMA es distinta a la propuesta por Teles y Brito (2005) ya que modelan las series temporales de los centros y de los radios de los intervalos, en lugar de los mínimos y los máximos. El modelo híbrido que proponen se basa en el modelo desarrollado por Zhang (2003) para series temporales clásicas. Dicho modelo combina un modelo ARMA y un perceptrón multicapa para obtener las predicciones.

En los últimos años, han surgido varios estudios que han documentado métodos referentes a la modelización y pronóstico de ITS (Arroyo & Maté, 2006; Arroyo et al., 2011; Arroyo, Roque, Maté, & Sarabia, 2007; González-Rivera & Lin, 2013; Hanet et al., 2015; He & Hu, 2009; Maia & Carvalho, 2011; Maia et al., 2008; Neto & Carvalho, 2008, 2010). Arroyo et al. (2007) propusieron el método de suavizado exponencial y lo compararon con otros enfoques para el modelado de series. Consideraron que la forma más directa de transformar una serie simbólica de intervalo es por un par de series de tiempo clásica. González-Rivera and Lin (2013) propusieron un modelo de regresión restringido para los bordes de la ITS que preserva el orden natural del intervalo en todos los casos. La idea fundamental de estos enfoques antes mencionados para modelar una ITS univariada es representarla como un vector, esto es, el punto medio y el rango, o ambos bordes de la serie y luego hacer la predicción de estas series, basados en

herramientas estadísticas clásicas. Han et al. (2015) propusieron una nueva estrategia via una clase de modelos intervalo condicional autorregresivo (ACI) para una ITS univariada. A diferencia de otros, el enfoque ACI propone un modelo parsimonioso que considera una observación intervalar como una unidad inseparable. Yang and Han (2015) exploraron la potencia del pronóstico de un modelo ACI para los datos bursátiles diarios de EE. UU. Y mostraron que estos modelos producen pronósticos de volatilidad más precisos que los métodos basados en modelos de series de tiempo puntuales. Han et al. (2016) proponen un modelo parsimonioso IVARMA para investigar la dinámica de un sistema de vectores ITS y para pronosticar un vector ITS. El modelo IVARMA generaliza el modelo vectorial autorregresivo de promedios móviles clásico para una serie de tiempo vectorial a la configuración de intervalo. Esto se logra extendiendo el trabajo de Han et al. (2015) para una ITS univariada al caso de un sistema ITS multivariado. Para estimar los parámetros de un modelo IVARMA usando datos de intervalo, se propone un método de estimación de mínima distancia y se establece la teoría asintótica de los estimadores propuestos.

3. RESULTADOS OBTENIDOS/ESPERADOS

4.

En el contexto del mencionado proyecto, en Octubre del año 2021, se presentó la comunicación *Pronóstico de la tasa de natalidad haciendo uso del Análisis de Datos Simbólicos*, en el XIV Congreso Latinoamericano de Sociedades Estadísticas. En este trabajo, considerando como fuente el Instituto Nacional de Estadísticas y Censos (INDEC) y la Dirección de Estadísticas e Información de Salud (DEIS), se realizó

una predicción de la tasa de natalidad en Argentina, utilizando una adaptación del método de Alisado Exponencial Simple. Para esto se emplearon paquetes del Software R y algoritmos desarrollados en Octave (versión libre de Matlab).

Por otra parte, la Lic. Cecilia Martínez trabajará en su tesis de maestría pronósticos de ITS simulados y con datos reales, empleando el método IVARMA y métodos de suavizado exponencial, propuestos en la literatura.

5. FORMACIÓN DE RECURSOS HUMANOS

El equipo de investigación del proyecto mencionado en la introducción está formado por docentes investigadores de dos facultades de la UNSJ, algunos de ellos son jóvenes investigadores. En particular, la Lic. Cecilia Martínez y Lic. Andrea Salas han investigado durante los dos primeros años acerca de los métodos desarrollados, en el contexto de ADS, para el pronóstico de ITS. La primera desarrolla su tesis de Maestría en Matemática, de la Universidad Nacional de San Luis, en esta línea. La segunda es alumna del Doctorado en Estadística de la Universidad Nacional de Rosario.

6. BIBLIOGRAFÍA

Arroyo, J., Espínola, R., & Maté, C. (2011). *Different approaches to forecast interval time series: A comparison in finance*. Computational Economics, 37(2), 169–191.

Froelich, W., & Salmeron, J. L. (2014). *Evolutionary learning of fuzzy grey cognitive maps for the forecasting of multivariate, interval-valued time series*. International Journal of Approximate Reasoning, 55(6), 1319–1335.

García-Ascanio, C., & Maté, C. (2010). *Electric power demand forecasting using*

interval time series: A comparison between VAR and iMLP. Energy Policy, 38(2), 715–725.

Golan, A., & Ullah, A. (2015). *Interval estimation: An infometrics approach*. Working Paper.

González-Rivera, G., & Lin, W. (2013). *Constrained regression for interval-valued data*. Journal of Business & Economic Statistics, 4, 473–490.

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton university press.

Han, A., Hong, Y., & Wang, S. (2016). *A Vector Autoregressive Moving Average Model for Interval-Valued Time Series Data*.

Published online: 23 Jun 2016; 417-460.

Han, A., Hong, Y., & Wang, S. (2015). *Autoregressive conditional models for interval-valued time series data*, Working Paper.

Han, A., Lai, K. K., Wang, S., & Xu, S. (2012). *An interval method for studying the relationship between the Australian dollar exchange rate and the gold price*. Journal of Systems Science and Complexity, 25(1), 121–132.

Maia, A. L. S., & Carvalho, F.A.T. (2011). *Holt's exponential smoothing and neural network models for forecasting interval-valued time series*. International Journal of Forecasting, 27(3), 740–759.

Nakajima, J. (2011). *Time-varying parameter VAR model with stochastic volatility: An overview of methodology and empirical applications*. Institute for Monetary and Economic Studies, Bank of Japan.

Teles, P. y Brito, M. P. (2005) *Modelling interval time series data*. En Proceedings of the 3rd IASC World Conference on Computational Statistics & Data Analysis. Limassol, Cyprus.

Yang, W., & Han, A. (2015). *A new approach for forecasting the price range with financial intervalvalued time series*

data. ASCE-ASME Journal of Risk
Uncertainty Engineering Systems, Part B:
Mechanical Engineering, 1(2), 021004,
1–8. doi:10.1115/1.4028751

Administración y Recuperación de Datos Multimedia Masivos

Luis Britos, Fernando Kasián, Verónica Ludueña, Franco Merenda, Diego Olivares,
Marcela Printista, Nora Reyes, Patricia Roggero, Pablo Samat

LIDIC, Dpto. de Informática, Fac. de Cs. Físico Matemáticas y Naturales, Universidad Nacional de San Luis
{fkasian, vlud, mprinti, nreyes, proggero}@unsl.edu.ar,
{ldoorz, merenda.franco83, olivarestello.diego, samatpablo}@gmail.com

Karina Figueroa

Universidad Michoacana de San Nicolás de Hidalgo, México
karina@fismat.umich.mx

Claudia Deco

Facultad de Ciencias Exactas, Ingeniería y Agrimensura, Universidad Nacional de Rosario
deco@fceia.unr.edu.ar

Resumen

Dado el crecimiento sostenido en los últimos tiempos de la disponibilidad de diferentes tipos de datos multimedia, por el uso masivo de dispositivos que producen datos digitalizados, se ha vuelto necesario contar con grandes repositorios de datos no estructurados provenientes de distintas fuentes, los cuales son difíciles de administrar bajo el modelo relacional de base de datos. Este crecimiento sostenido tanto de la cantidad de datos en formato digital disponible, como en la variedad de los mismos, ha ido acompañado por la rápida evolución de las tecnologías de la información y comunicación. Al considerar tipos de datos tales como texto libre, imágenes, audio, video, secuencias biológicas de ADN o proteínas, entre otros, su administración y las distintas formas de recuperación de datos no se corresponden a las habituales y forzar una modelización de los mismos no adecuada podría restringir de antemano el tipo de operaciones que se puedan realizar sobre ellos.

Así, al considerar la administración y recuperación sobre grandes conjuntos de datos no estructurados, se hace evidente la necesidad de considerar nuevos modelos y herramientas para su administración y su indexación, considerando las operaciones de interés. Además, como el objetivo de cualquier sistema de recuperación de información es obtener información útil para el usuario, mediante consultas a la base de datos, no sólo se deben soportar distintos tipos de consultas sino también resolverlas de manera eficiente usando índices apropiados.

Palabras Claves: bases de datos masivas, computación de alto desempeño, recuperación de información.

1. Contexto

Esta línea de investigación “Recuperación de Datos e Información” se encuentra enmarcada dentro del Proyecto Consolidado 3-03-2018 de la Universidad Nacional de San Luis (UNSL - Res. CS 126/18)

y en el Programa de Incentivos (Código 22/F834): “Tecnologías Avanzadas Aplicadas al Procesamiento de Datos Masivos”, se desarrolla en el Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC) de la UNSL y su finalización, debido a la pandemia, se postergó a 2023.

El principal objetivo de esta línea es desarrollar herramientas eficientes para administrar bases de datos masivas, que almacenan datos no estructurados. En este sentido, se analizan nuevas técnicas que provean una buena interacción con el usuario y se desarrollan nuevas estructuras de datos capaces de indexar un gran volumen de datos no estructurados, que permitan resolver eficientemente consultas de interés para la recuperación de información sobre estos tipos de datos. Para ello, se considera el diseño y desarrollo de índices para conjuntos de datos no estructurados masivos (datos multimedia, texto, secuencias de ADN, huellas digitales, etc.), que sean eficientes y escalables, para memorias jerárquicas, aplicando técnicas de computación de alto desempeño (HPC). Además, se busca su incorporación en un sistema de administración para estas bases de datos, que apoye la recuperación de información sobre datos multimedia masivos.

2. Introducción y Motivación

Debido al uso masivo de internet, a la gran disponibilidad de dispositivos electrónicos y a la evolución de las tecnologías de información y comunicación, se están generando una cantidad masiva de datos. Además, por provenir de fuentes muy diversas, los tipos de datos generados son también muy variados. En este contexto de problemas de “big data”

(porque aparecen sus dos características importantes, que son: el volumen de datos y variedad de los mismos) se hace imposible restringir las consultas a búsquedas sobre datos estructurados tradicionales, porque obligaría a representar una visión parcial del problema. Por ello, para no perder información relevante, se deben redefinir las técnicas de procesamiento, análisis y obtención de información útil y formular nuevas metodologías.

Un enfoque útil para sistemas de recuperación de información es *la búsqueda basada en contenidos*, donde se usa el dato no estructurado en sí mismo para describir lo que se está buscando. En este escenario son aplicables las búsquedas por similitud, por ser más generales y aplicables. Para lograr resolver dichas búsquedas eficientemente, se deben considerar *métodos de acceso* o *índices métricos* [3].

Por lo tanto, al considerar grandes cantidades de datos no estructurados y ser necesario responder consultas de recuperación de información, se pueden utilizar estos índices métricos para obtener las respuestas de manera más eficiente. En aplicaciones reales, estos índices en general además de eficientes, deben permitir actualizaciones (altas y bajas de elementos) y ser escalables; es decir, no degradar su desempeño significativamente a medida que crece el volumen de datos a indexar.

Los *espacios métricos*, compuestos por un *universo* \mathcal{U} de objetos y una *función de distancia* $d : \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}^+$, resultan adecuados para modelizar las búsquedas por similitud. En particular, d cumple con las propiedades de una métrica (reflexividad, positividad estricta, simetría y desigualdad triangular). La función de distancia permite medir la disimilitud entre dos objetos. Es habitual considerar dos tipos básicos de búsqueda por similitud: *la búsqueda por rango* y *la búsqueda de los k vecinos más cercanos*. El cálculo de distancia sobre algunos tipos de datos no estructurados puede ser muy costoso. Por lo tanto, un objetivo importante en el diseño de los índices es ahorrar cálculos de distancia. Una de las principales ventajas de este modelo es que, además de brindar un marco formal, es independiente del dominio de la aplicación. Al considerar una base de datos $S \subseteq \mathcal{U}$ tal que $|\mathcal{S}| = n$, cualquier consulta se resuelve de manera trivial calculando n evaluaciones de distancia. Las distancias en general son costosas de calcular (por ej.: comparación de huellas digitales). Por ello, en la mayoría de las aplicaciones sobre datos masivos no basta con resolver el problema de manera trivial. Así, para minimizar los cálculos de

distancia, se preprocesa S para construir un índice.

Sin embargo, en muchos casos es probable que la base de datos, el índice, o ambos, no puedan almacenarse en memoria principal y deban alojarse en memoria secundaria. Esto puede incrementar mucho los costos de las búsquedas, dado que las operaciones de E/S en memoria secundaria suelen ser muy costosas. Por lo tanto, para lograr eficiencia, no basta con realizar pocos cálculos de d sino que también se debe minimizar el número de operaciones de E/S. Para ello, hay que considerar el nivel de la jerarquía de memorias sobre la que se trabaja, utilizar técnicas de computación de alto desempeño (HPC) y, en algunos casos, admitir respuestas no exactas.

En esta línea, se busca desarrollar nuevas técnicas y aplicaciones que soporten la interacción con el usuario, diseñar índices que permitan la manipulación eficiente de grandes volúmenes de datos no estructurados y faciliten la realización de diferentes tipos de consultas, para construir herramientas de recuperación de información sobre conjuntos masivos de datos. De esta manera, se espera contribuir al desarrollo de soluciones a problemas de big data para aplicaciones reales.

3. Líneas de Investigación

Dado que en esta línea se pretende contribuir a distintos aspectos de los sistemas de recuperación de información sobre grandes volúmenes de datos no estructurados, se ha considerado proveer de un administrador de sistemas para este tipo de bases de datos y, por otra parte, optimizar índices existentes, diseñar nuevos índices, resolver distintas clases de consultas, y lograr eficiencia y escalabilidad para grandes volúmenes de datos.

DBMS para Bases de Datos Multimedia

A pesar de que las operaciones más comunes sobre bases de datos multimedia son las búsquedas por rango o de k -vecinos más cercanos, la operación de *join* por similitud debería brindarse en un sistema administrador para bases de datos multimedia [10].

Hay distintas variantes del *join* por similitud entre dos bases de datos A y B , con $A, B \subseteq \mathcal{U}$, dependiendo del criterio de similitud Φ utilizado. El resultado de cualquiera de las variantes de esta operación es un conjunto de pares formados (x, y) , $x \in A$ e $y \in B$, tales que se satisface el criterio de similitud Φ entre x e y . Las variantes más conocidas son: el *join* por rango, el *join* de k -vecinos más cercanos y el *join* de k pares de vecinos más cercanos; entre

otras. Formalmente, dadas $A, B \subseteq \mathcal{U}$, el *join por similitud* entre A y B es $(A \bowtie_{\Phi} B) = \{(x, y) / x \in A \wedge y \in B \wedge \Phi(x, y)\}$. Si $A = B$, la operación se denomina *auto-join*. Las variantes más conocidas son: el join por rango, el join de k -vecinos más cercanos y el join de k pares de vecinos más cercanos; entre otras. Al resolver el join por similitud es posible que ambas, una o ninguna de la bases de datos posean un índice; o que ambas bases de datos se indexen conjuntamente con un índice diseñado para el join [9]. Por lo tanto, un DBMS para estas bases de datos debe considerar cómo resolver eficientemente las consultas en función del índice utilizado.

PostgreSQL es un DBMS que permite realizar consultas por similitud sobre algunos atributos. Por ejemplo, si se han creado índices *KNN-GiST* sobre texto, se podrían resolver búsquedas de k -vecinos más cercanos sobre ellos. Sin embargo, estos índices brindan plantillas sólo para crear *árboles balanceados* tales como: *B-tree* o *R-tree*. Pero, se ha mostrado que el “balance” no siempre es bueno para los índices sobre espacios métricos [1]. Por otro lado, *PostgreSQL* no dispone de este tipo de consultas sobre todo tipo de datos métricos. Por lo tanto, es importante incorporar en un DBMS la capacidad de manejar distintos tipos de datos no estructurados y todas las operaciones de interés sobre ellos.

En las respuestas a consultas de join por similitud pueden existir pares muy similares entre sí. Por lo tanto, para acelerar los tiempos de respuesta, disminuyendo los tamaños de los conjuntos obtenidos, se pueden aplicar técnicas de diversificación de las respuestas [11]. En este caso, se obtiene un conjunto más pequeño de pares, con respuestas útiles y diversificadas, y en menor tiempo. Otra manera de acelerar el tiempo de respuesta es brindando una respuesta aproximada a la consulta. Por ello, se ha propuesto un algoritmo simple y eficiente que permite responder consultas de join por similitud de k vecinos más cercanos aproximados dentro del mismo conjunto (auto-join), con una razonable precisión en la respuesta [4]. Por lo tanto, se espera incorporar éste y otros de los desarrollos de esta línea a *PostgreSQL* para obtener un DBMS más aplicable en sistemas de información sobre datos multimedia masivos.

Métodos de Acceso Métricos

Como se ha mencionado, los métodos de acceso métricos (MAMs) permiten responder a las búsquedas sobre conjuntos de datos no estructurados de manera más eficiente que si se examina toda la base

de datos [3]. Los MAMs aprovechan las distancias almacenadas en el índice y que la función de distancia satisface la desigualdad triangular para ahorrar cálculos de distancia y tiempo. En general, los MAMs mantienen algunas distancias precalculadas entre los elementos de la base de datos y objetos particulares o distinguidos. Así, durante las búsquedas se puede estimar la distancia entre cualquier objeto de consulta q y los elementos de la base de datos utilizando la desigualdad triangular, sin calcular d realmente. Existen dos enfoques principales para los MAMs, dependiendo si esos objetos distinguidos son *pivotes* o *centros*. Si los elementos distinguidos son pivotes, se almacenan las distancias de todos los objetos de la base de datos a ellos. Por el contrario, si son centros se particiona el espacio en zonas denominadas *particiones compactas*, por cercanía a los centros. Estas particiones almacenan un radio de cobertura que determina la zona de cada centro.

Cuando se diseñan MAMs, se deben considerar varios aspectos: dinamismo (que soporten altas y bajas), en qué nivel de la jerarquía de memorias deben almacenarse, si es posible aplicar técnicas de HPC para mejorar los tiempos de respuesta, si son de respuesta exacta o de respuesta aproximada y la dimensionalidad del espacio métrico considerado.

Dinamismo: Si la base de datos se conoce de antemano, se construye el índice (*estático*) y luego se realizan las consultas. Por el contrario, la única manera de construir el índice (*dinámico*) es a medida que se incorporan los elementos; es decir, de manera incremental, se considera que las búsquedas pueden realizarse en cualquier momento. Los índices estáticos pueden seleccionar los mejores objetos distinguidos para el índice, los dinámicos no.

Jerarquía de Memorias: Otro aspecto importante para buscar una solución es saber si se puede trabajar en memoria principal o, por el contrario, si por ser conjuntos de datos masivos se deberá trabajar en otros niveles de la jerarquía de memorias. En caso que el índice deba alojarse en memoria secundaria, se deben minimizar la cantidad de cálculos de distancia y también el número de operaciones de E/S.

Computación de Alto Desempeño: En algunos casos, si no se logra la eficiencia deseada mediante la optimización del índice en sí mismo, se pueden aplicar técnicas de computación de alto desempeño (HPC) para acelerar el tiempo de respuesta.

Exactitud de la Respuesta: Otra manera de acelerar la respuesta a una consulta por similitud es admitir una respuesta aproximada, permitiendo que la misma no sea exacta, pero si muy rápida.

Por el volumen de los datos con los que se trabaja (por ejemplo, millones de imágenes en la Web) se debe considerar que posiblemente los índices deban almacenarse en otros niveles de la jerarquía de memoria. En general, se considera directamente su almacenamiento en memoria secundaria y sólo disponer de memoria principal para mantener información muy útil que permita determinar que partes del índice deben recuperarse desde el disco. Por ello, para lograr eficiencia, el índice debe minimizar tanto el número de cálculos de distancia como de operaciones sobre el disco (E/S) en las búsquedas.

Así, una parte de esta línea se dedica a diseñar MAMs, preferentemente dinámicos y adaptados para memoria secundaria, cuyo desempeño en las búsquedas sea adecuado para las aplicaciones de interés. Algunos de ellos se basan en la *Lista de Clusters (LC)* [3, 2] son totalmente dinámicos, es decir, admiten inserciones y eliminaciones de objetos y están especialmente diseñados para trabajar sobre grandes volúmenes de datos [8]. Uno de ellos, el *Conjunto Dinámico de Clusters (DSC)*, que tiene una buena ocupación de página y operaciones eficientes tanto en cálculos de distancia como en operaciones de E/S, logra un buen desempeño en espacios de alta dimensión. El *DSC* mantiene los clusters en memoria secundaria y organiza los centros de clusters en un *DSAT* en memoria principal, para lograr búsquedas con menos cálculos de distancia y accesos páginas/clusters. La información en el *DSAT* permite mejorar los costos en cálculos de distancia, y mantener bajos los costos de acceso a disco.

Sin embargo, cada inserción en un *DSC* debe localizar el clúster donde se insertará el nuevo elemento. Luego, se realiza una lectura del clúster desde el disco y se verifica si el nuevo objeto podría ser un mejor centro para el clúster, haciendo si es posible el clúster más compacto. Finalmente, se graba nuevamente el clúster en el disco, actualizando además el *DSAT* de centros si el nuevo objeto se transformó en el nuevo centro. Aunque esta operación, además de los cálculos de distancia necesarios, realiza dos operaciones de E/S en el disco (podrían ser 3 si el clúster estaba ya lleno), es posible considerar bajar esta cantidad de operaciones de E/S. Para ello, se está analizando una variante para el *DSC* en la cual, en lugar de insertar los elementos en el índice

a medida que van llegando, se demora la incorporación de cada nuevo objeto a un clúster, hasta tener varios elementos y poder determinar un mejor agrupamiento de los mismos, que mejore los costos de búsqueda, al lograr clusters aún más compactos y que aseguren una total ocupación de la página del disco, achicando el tamaño del archivo y reduciendo los tiempos de acceso. Esta opción, además podría reducir significativamente el costo de construcción, por amortizar el costo de la escritura de un clúster en disco, luego de varias inserciones.

Así, se ha diseñado el *Buffered On Line Dynamic Set of Clusters (BOLDSC)*, una variante del *DSC*, que agrega un espacio en memoria para mantener los objetos que son insertados (bolsa), además de un índice auxiliar (de pivotes) sobre estos elementos. Como pivotes se eligen elementos de la bolsa, y cuando ésta se llena, se selecciona el pivote que necesita el menor radio de cobertura para encerrar la cantidad de elementos que caben en un clúster y todos éstos se sacan de la bolsa y se graban en un clúster completo. La grabación de clusters llenos mejora la cantidad de E/S, pero puede provocar que objetos cercanos, que se insertan en diferentes momentos, queden en clusters diferentes, ocasionando degradación en las búsquedas. Por ello, se está estudiando una variante de la BOLDSC que graba clusters con una cantidad de objetos que está en función de un factor de carga ρ (casi llenos). Como los clusters no estarán llenos, ellos tendrán un espacio libre en cada clúster. Esto se considera al insertar un nuevo objeto. Primero se verifica si hay clusters que lo pueden incorporar y de ellos se elige al más cercano a él. Se verifica también si el nuevo elemento sería un mejor centro para el clúster y si es así, se actualiza el *DSAT* de centros. Si no es posible incorporarlo a un clúster (todos llenos), se lo incorpora a la bolsa y se actualiza el índice de pivotes. Así, los clusters se adaptan y se mejoran los costos en las búsquedas.

El *DSC* basa su buen desempeño en la calidad de la partición generada por los centros de sus clusters. Por lo tanto, otro aspecto a considerar es mejorar la poda en estos agrupamientos. Cada clúster en *DSC* está determinado por su centro y su radio de clúster. Sin embargo, puede haber “huecos” dentro de ellos. Por esta razón, es posible seleccionar un conjunto de “pivotes” globales que permitan caracterizar sobre cada uno de los clusters las zonas dentro de las cuales existan efectivamente elementos [7]. De esta manera, se puede mantener para cada clúster información sobre la mínima y la máxima distancia

a la que cada pivote encuentra elementos de dicho clúster. Esta información se utiliza para identificar los “huecos” del clúster para que luego, durante las consultas, se pueda descartar un clúster si la bola de la consulta no posee intersección efectiva con la zona del clúster que realmente contiene elementos de interés. En las búsquedas se recupera desde memoria secundaria cada clúster que intersecta la bola de consulta con centro q y radio r . Al identificar las zonas del clúster que no contienen efectivamente elementos, si la bola de consulta intersecta al clúster en su zona “hueca” se puede descartar el clúster y ahorrar cálculos de distancia y lecturas a disco.

Otra manera de acelerar las búsquedas es paralelizar el índice usando técnicas de HPC. Así, se está trabajando en el diseño e implementación de una versión paralela del *Conjunto Dinámico de Clusters* (DSC) [8]; esperando, no sólo bajar la cantidad de cálculos de distancia y operaciones de E/S necesarias para responder a una consulta, sino aumentar la cantidad de consultas simultáneas a resolver, aprovechando al máximo las operaciones de E/S que se realicen durante las mismas. Para ello, se aplicarán y compararán distintas versiones paralelas del DSC.

4. Resultados

Recientemente, además de una estrategia para obtener buenas listas de candidatos para las búsquedas por similitud usando permutaciones en un índice invertido [5], se ha publicado una manera de acortar dichas listas de candidatos para acelerar las búsquedas [6]. Por otro lado, se está evaluando experimentalmente la versión paralela del índice DSC, diseñada para memoria secundaria, que admite inserciones/eliminaciones de objetos y permitirá responder eficientemente lotes de consultas. Además, se encuentran en etapa de evaluación la *BOLDSC* y en etapa de implementación las otras mejoras al DSC. Se continúa trabajando en la extensión de *PostgreSQL*, probando el prototipo de DBMS para consultas por similitud sobre distintos tipos de datos.

5. Formación de Recursos

Se están realizando las tesis de Maestría: (1) - “Estructuras Eficientes sobre Datos Masivos para Búsquedas en Espacios (UNSL)”, (2) - “Sistema Administrador para Bases de Datos Métricas” (UNSL), (3) - “Índices Métricos – Optimización del DSC usando Cortes de Regiones” (UNSJ) y (4) - “Optimización del BOLDLC por la Mejora de la Densidad y Solapamiento de los Clusters”(UNSJ).

Referencias

- [1] E. Chávez, V. Ludueña, and N. Reyes. Revisiting the VP-forest: Unbalance to improve the performance. In *Proc. de las JCC08*, page 26, 2008.
- [2] E. Chávez and G. Navarro. A compact space decomposition for effective metric indexing. *Pattern Recognition Letters*, 26(9): 1363–1376, 2005.
- [3] E. Chávez, G. Navarro, R. Baeza-Yates, and J. Marroquín. Searching in metric spaces. *ACM*, 33(3):273–321, 2001.
- [4] S. Ferrada, B. Bustos, and N. Reyes. An efficient algorithm for approximated self-similarity joins in metric spaces. *Information Systems*, 91:101510, 2020.
- [5] K. Figueroa, N. Reyes, and A. Camarena-Ibarrola. Candidate list obtained from metric inverted index for similarity searching. In *Advances in Computational Intelligence*, 29–38, 2020. Springer International Publishing.
- [6] K. Figueroa, A. Camarena-Ibarrola, and N. Reyes. Shortening the candidate list for similarity searching using inverted index. In *Pattern Recognition*, 89–97, 2021. Springer International Publishing.
- [7] J. Lokoč, J. Moško, P. Čech, and T. Skopal. On indexing metric spaces using cut-regions. *Information Systems*, 43:1–19, 2014.
- [8] G. Navarro and N. Reyes. New dynamic metric indices for secondary memory. *Information Systems*, 59:48 – 78, 2016.
- [9] R. Paredes and N. Reyes. Solving similarity joins and range queries in metric spaces with the list of twin clusters. *JDA*, 7:18–35, 2009.
- [10] C. Rong, C. Lin, Y. N. Silva, J. Wang, W. Lu, and X. Du. Fast and scalable distributed set similarity joins for big data analytics. In *IEEE 33rd International Conference on Data Engineering (ICDE)*, 1059–1070, 2017.
- [11] L. Santos, L. Carvalho, W. Oliveira, A. Traina, and C. Jr. Traina. Diversity in similarity joins. In *Similarity Search and Applications*, vol. 9371, LNCS, 42–53. Springer International Publishing, 2015.

CGIV – Computación Gráfica, Imágenes y Visualización

Índices y Operaciones para Bases de Datos Métricas

M. D. Alba, J. Arroyuelo, M. E. Di Genaro, A. Grosso, M. Jofré, V. Ludueña, N. Reyes
Dpto. de Informática, Fac. de Cs. Físico-Matemáticas y Naturales, Universidad Nacional de San Luis
{bjarroyu, digeme, agrosso, vlud, nreyes}@unsl.edu.ar, {mdaniela.alba, monicajofre}@gmail.com

Edgar Chávez

Centro de Investigación Científica y de Educación Superior de Ensenada, México

elchavez@cicese.mx

Karina Figueroa

Fac. de Cs. Físico-Matemáticas, Universidad Michoacana de San Nicolás de Hidalgo, México

karina@fisimat.umich.mx

Rodrigo Paredes

Dpto. de Cs. de la Computación, Fac. de Ingeniería, Universidad de Talca, Chile

raparede@utalca.cl

Resumen

La disponibilidad de dispositivos electrónicos en diversos ámbitos y al alcance de todos, junto con el uso masivo de las redes, ha provocado que una gran cantidad y variedad de datos sean generados cada segundo. Este contexto motivó que las bases de datos debieran adaptarse a nuevos tipos de datos (no estructurados) y evolucionaran para lograr administrar de manera eficiente ese gran volumen de datos, al igual que el tipo de requerimientos al que son sometidos los mismos.

Un modelo de base de datos que se adapta al entorno descrito son las *Bases de Datos Métricas*. Esta investigación pretende contribuir a la madurez de este modelo de bases de datos, considerando distintas perspectivas como la administración del espacio disponible (crucial debido a la gran cantidad de datos); formas más sofisticadas de búsqueda sobre las mismas; optimización de estos depósitos, o desarrollo de nuevos, considerando incluso la arquitectura del procesador, entre otros.

Palabras Claves: bases de datos métricas, índices, búsquedas por similitud.

Contexto

Esta investigación se realiza en el marco del Proyecto Consolidado *Tecnologías Avanzadas de Bases de Datos* de la Universidad Nacional de San Luis (Código 03-2218) y en Programa de Incentivos (Código 22-F814), dentro de la línea *Bases de Datos no Convencionales*. Colaboran investigadores de otros grupos de la región: Universidad de Talca (Chile), Universidad Michoacana de San Nicolás de

Hidalgo y Centro de Investigación Científica y de Educación Superior de Ensenada (México). Mediante RCS N° 186/2020 se prorrogaron los proyectos de la UNSL hasta diciembre de 2022.

Esta línea tiene como principal objetivo lograr la consolidación del modelo de Bases de Datos Métricas para soportar distintos tipos de bases de datos no convencionales. Para este propósito, se está investigando sobre obtener índices que sean escalables, capaces de manejar grandes volúmenes de datos, sin degradar significativamente su desempeño; con operaciones de E/S eficientes; y que además de ser dinámicos, permitan tanto inserciones como eliminaciones, resulten más eficaces considerando el nivel de la jerarquía de memorias en el que trabajen. Igualmente, se analizan nuevas arquitecturas del procesador en un intento de obtener mejoras, a un muy bajo nivel, en los administradores de estas bases de datos. Se espera, de esta manera, contribuir en diferentes campos de aplicación: sistemas de información geográfica, robótica, visión artificial, diseño asistido por computadora, computación móvil; entre otros.

Introducción

El acceso generalizado a dispositivos capaces de generar datos digitales en ámbitos tan diferentes como el educativo, productivo, laboral, de la salud, recreativo, científico, etc. ha ocasionado que el volumen de datos generados y almacenados y la variedad

de sus tipos de datos, crezca de manera exponencial. Como respuesta, las bases de datos han debido adaptarse rápidamente tanto a la cantidad de datos, que deben ser administrados eficientemente, como a su variedad y disimilitud. En este contexto los repositorios especializados en datos *no estructurados* se vuelven indispensables.

El tipo de requerimientos al que son sometidos estos datos también es variado, dado que los mismos pueden provenir de entornos muy diferentes; por ejemplo, se puede ingresar una melodía esperando encontrar canciones semejantes a dicho trozo, o buscar las huellas digitales más similares a una dada; en estos casos las búsquedas tradicionales (exactas) carecen de sentido, en cambio las *búsquedas por similitud* resultan más apropiadas. En este tipo de búsqueda se suele dar un objeto como modelo de lo que se quiere recuperar y se busca en la base de datos los objetos que sean suficientemente similares a él. Estas búsquedas se las conoce como consultas por contenido o consultas mediante un ejemplo (query by example).

Un modelo que resulta adecuado, por englobar muchas de las características compartidas por estas necesidades tan diversas, es el de *espacios métricos*. Este modelo es determinado por un universo de objetos y una función de distancia definida entre ellos, que mide cuán diferentes son. Cualquier tipo de objetos no estructurados, que admita la definición de una medida que para un par de elementos indique cuán diferentes son los mismos, admite ser modelizado de esta manera. La única restricción es que esa medida cumpla con las propiedades que la hagan una métrica. En general, esas medidas sobre conjuntos particulares de datos son provistas por expertos (por ejemplo: distancias que sirven para comparar huellas dactilares).

Además, para responder eficientemente a requerimientos tan dispares sobre estas bases de datos, sin realizar una examinación secuencial del conjunto de datos, son necesarios los llamados *Índices Métricos* o *Métodos de Acceso Métricos* (MAMs). La actualización y optimización de estos índices se vuelve esencial cuando se considera la variedad de ámbitos en los que se aplica este modelo (reconocimiento de voz, reconocimiento facial, bases de datos médicas, minería de datos, biología computacional, etc.), para lograr su adaptación a cada caso particular y afrontar retos como el soporte de conjuntos masivos de datos, el permitir actualizaciones (inserciones/eliminaciones), la resolución de búsquedas más

complejas y mejorar el desempeño de los administradores de bases de datos (DBMS) también a bajo nivel.

Líneas de Investigación y Desarrollo

Bases de Datos no Convencionales

En este ámbito de investigación, las bases de datos que administran vídeos, imágenes, texto libre, secuencias de ADN o de proteínas, audio, etc., las llamadas *bases de datos no convencionales*, serán modelizadas utilizando el modelo de espacios métricos, por lo que también son referenciadas como *bases de datos métricas*.

Para responder eficientemente a los distintos requerimientos sobre las mismas, evitando comparaciones exhaustivas sobre la base de datos durante una consulta por similitud, se requiere el uso de índices especializados. Por ello, un objetivo prioritario es optimizar los mismos, analizando aquellos que han mostrado buen desempeño en las búsquedas para reducir su complejidad considerando, cuando sea necesario, el nivel de la memoria en lo que se lo alojará y/o priorizando además, cuando sea posible, su dinamismo y escalabilidad. Debido a lo costoso que suele resultar calcular la distancia entre dos objetos, el número de cálculos realizados al crear el índice o al realizar búsquedas, es usado como medida general de complejidad y será el parámetro que se debe optimizar.

Formalmente, un espacio métrico está definido por un universo de objetos \mathbb{U} y una función de distancia entre ellos $d : \mathbb{U} \times \mathbb{U} \mapsto \mathbb{R}^+$, que permite medir la disimilitud entre los objetos del universo. En particular, d cumple con las propiedades de una métrica (reflexividad, positividad estricta, simetría y desigualdad triangular), lo que resulta muy útil al momento de resolver consultas por similitud. En particular, la propiedades que permiten ahorrar cálculos de distancia son la simetría y la desigualdad triangular. Una base de datos X es un conjunto $X \subseteq \mathbb{U}$ y una consulta por similitud sobre la misma, $q \in \mathbb{U}$, en general suele ser de dos tipos: *por rango* (q, r) o de *k-vecinos más cercanos* (k -NN(q)).

Búsqueda de los k Vecinos

El localizar las estaciones de servicio más cercanas a nuestra ubicación, es un requerimiento que puede resolverse en espacios métricos mediante una consulta k -NN. Este tipo de consulta resulta muy útil

en una variedad de aplicaciones como la predicción de funciones, el aprendizaje automático, la cuantificación y compresión de imágenes, etc. Una variación menos utilizada de la misma es la que permite obtener los k -vecinos más cercanos de *todos* los elementos de una base de datos, la *All- k -NN*, que resulta tan útil como la anterior. El planteo general de esta consulta relaciona cada elemento $u \in X$, con los k objetos en $X - \{u\}$ que tengan la menor distancia a él.

La solución ingenua tiene una complejidad de $O(n^2)$ cálculos de distancia, si $|X| = n$, y consiste en comparar cada objeto en la base de datos con todos los demás, para obtener la respuesta a esta consulta. Evidentemente, es necesario proceder de manera mucho más eficiente, por ejemplo a través de un preprocesamiento la base de datos; así se construye un índice, y luego se buscan en él los k -NN de cada elemento del conjunto.

Sin embargo, en algunas situaciones esta solución puede resultar tan ineficiente como la ingenua. Esto ocurre cuando se trabaja con espacios métricos de alta dimensión, o si la función de distancia utilizada es demasiado costosa de calcular, o cuando se administra una base de datos masiva. Estas circunstancias pueden requerir revisar la base de datos completa, sin importar la estrategia implementada. Por otro lado, al considerar los requerimientos de algunas aplicaciones particulares, hay algunos que priorizan la velocidad de respuesta sobre la precisión de la misma [12, 6, 13, 7]. En este contexto, toman preponderancia las llamadas *búsquedas por similitud aproximadas*, que admiten mejorar la velocidad de la respuesta a las consultas a costa de aceptar algunos “errores” en la respuesta.

El *Grafo de los k -vecinos más cercanos (k NNG)* [11] se encuentra entre las soluciones propuestas para espacios métricos generales. Este grafo asocia cada elemento de la base de datos a sus k vecinos más cercanos y resulta ser un índice eficiente, con respecto a algunas de las técnicas clásicas. Sabiendo que resolver el problema de los *All- k -NN* permite construir el k NNG, se han propuesto nuevas técnicas, que responden al problema de *All- k -NN*, y permiten computar una aproximación del k NNG. Éstas, conectan cada objeto u de la base de datos con k vecinos *cercanos*, relajando la condición que exige que no haya, en toda la base de datos, algún objeto más cercano a u que los k vecinos devueltos. Aunque estas técnicas pueden perder algún objeto muy cercano a u y en su lugar devolver otro un poco más

lejano obtienen, a cambio, una respuesta más rápida. Al grafo construido de esta manera se lo denomina *Grafo de vecinos cercanos (kn NG)* [4]. Una característica común en estas propuestas es que ninguna resuelve el problema a través de *buscar en un índice*.

El profundo conocimiento que se tiene del *Árbol de Aproximación Espacial DistalDiSAT* [5] permite plantear una primera aproximación que propone un enfoque ingenioso para un caso particular del problema, se resuelve el *All-1-NN*. Esta técnica utiliza información provista por la *construcción del DiSAT* para obtener el $1n$ NG, conectando a cada objeto con un elemento *cercano* de la base de datos, que puede ser, o no, su vecino más cercano [4]. Esta propuesta permite recuperar el $1n$ NG con bajo costo, logrando muy buena precisión, un error bajo y por ende un buen compromiso calidad/tiempo, y *sin realizar ninguna búsqueda en el índice*.

Las demás propuestas desarrolladas no se apoyan en ningún índice y se enfocan en el problema general, resolviendo la consulta *All- k -nN* y computando el kn NG. La base de estos desarrollos es aprovechar de manera ingeniosa las propiedades que cumple la *función de distancia*. En ellas se sugieren distintas maneras de seleccionar muestras de la base de datos, a partir de las cuales se obtiene un conjunto de distancias que serán el punto de partida de este proceso. Se analizan diferentes maneras de utilizar la información conseguida para calcular, en algunos casos, los vecinos exactos [3] y en otros los aproximados para todos los objetos de la base de datos, utilizando las propiedades mencionadas previamente de simetría o desigualdad triangular.

Índices Métricos

Como se ha explicado, los índices métricos o MAMs permiten responder a las búsquedas sobre conjuntos de datos no estructurados de manera mucho más eficiente que si se examina toda la base de datos de manera secuencial [6]. Para hacerlo los índices métricos aprovechan tanto las distancias almacenadas en él, cómo el hecho de que la función de distancia satisface la simetría y la desigualdad triangular, ahorrando de esta forma cálculos de distancia y tiempo.

Teniendo en cuenta el variado espectro de aplicación de los índices métricos, al momento de estudiar su optimización se deben considerar diferentes puntos de vistas; como tener en cuenta su dinamismo, analizar si se adaptan adecuadamente al almacenamiento en memoria secundaria, considerando como

medida de complejidad en ese caso no sólo el número de cálculos de distancia, sino también el número de operaciones de E/S necesarias, entre otros.

El desarrollo del *Árbol de Aproximación Espacial Dinámico (DSAT)* [10], logró proveer dinamismos a uno de los índices de mejor desempeño en espacios métricos de mediana a alta dimensión, el *Árbol de Aproximación Espacial (SAT)*, el cual a pesar de su eficiencia era totalmente estático. El *DSAT* permite realizar inserciones y eliminaciones; es decir, se puede crear incrementalmente, manteniendo el buen desempeño en las búsquedas. Sin embargo, el *DSAT* agrega un parámetro a sintonizar, mientras que el *SAT* carecía de parámetros.

Siguiendo con el enfoque del dinamismo, se propuso en primera instancia el desarrollo de la *Foresta de Aproximación Espacial Distal (DiSAF)* [2]. Este índice pretendía dinamizar el *Árbol de Aproximación Espacial Distal (DiSAF)*, que además de no necesitar ningún parámetro extra, lograba mejorar las búsquedas respecto de sus antecesores *SAT* y *DSAT*, sin embargo también era estático. La *DiSAF* es dinámica y aplica la técnica de dinamización de Bentley y Saxe al *DiSAF*. Además, aprovecha el profundo conocimiento que se tiene sobre la aproximación espacial para lograr mejorar al máximo su desempeño. Este índice se diseñó para memoria principal, sin embargo, los costos de construcción son altos debido a la necesidad de reconstruir subárboles luego de cada inserción. Como consecuencia del análisis de los motivos que hacían costosa a la *DiSAF*, principalmente en su construcción, se diseñó el *Árbol de Aproximación Espacial Distal Dinámico (DDiSAT)*. Esta versión utiliza *inserción perezosa* para amortizar los costos de reconstrucción, mientras se mantiene su desempeño en las búsquedas. En la actualidad se están evaluando sus resultados sobre diferentes bases de datos métricas.

Otro aspecto a considerar es cuando los índices no caben en memoria principal, ya sea porque administran una base de datos masiva, o porque los objetos almacenados en la misma son muy grandes (imágenes satelitales). Entonces surge la necesidad de diseñar índices que utilicen adecuadamente la memoria secundaria. En este contexto, se está trabajando en lograr una versión dinámica del *DDiSAT* para memoria secundaria, que además de amortizar los costos de reconstrucción entre varias inserciones y mantener un buen desempeño en las búsquedas, se adapte a memoria secundaria realizando un buen uso de las páginas de disco, a fin de minimizar el número

de operaciones de E/S. Para lograrlo, se debe considerar no sólo que logre un desempeño comparable al de la versión de memoria principal en cantidad de cálculos de distancia, sino que las páginas en disco mantengan una buena ocupación, que la cantidad de operaciones de E/S necesarias sea baja y que se mantenga la localidad en los accesos.

Los requerimientos de aplicaciones que priorizan la rapidez en las respuestas, a costa de perder algunos elementos de la misma, es una faceta tan importante como las consideradas anteriormente. A este tipo de búsquedas, en las que se intercambia precisión (devolviendo sólo algunos objetos relevantes) por velocidad en la respuesta (esos objetos se devuelven más rápido), se denominan *aproximadas*. Para conjuntos de datos masivos, las búsquedas por similitud aproximadas permiten obtener un buen balance entre el costo de las búsquedas y la calidad de la respuesta obtenida.

El diseño de la *Lista Dinámica de Permutaciones Agrupadas (DLCP)* [8], que además de ser dinámica es para memoria secundaria, fue diseñada usando como base uno de los mejores representantes de este tipo de consultas, el *Algoritmo Basado en Permutaciones (PBA)* [1], que logra una respuesta de alta calidad a un bajo costo. La *DLCP*, que combina *LC* con *PBA*, agrupa por distancia entre las *permutaciones* de los objetos, en lugar de hacerlo por distancia entre objetos, y además se le puede indicar cuántos cálculos de distancia y/o operaciones de E/S utilizar, para obtener una respuesta rápida, aunque menos precisa. Actualmente, se está considerando también una versión de *DiSAF* para memoria secundaria, que mantenga el dinamismo en la construcción y que permita aprovechar la información de los *DiSAF* que la integran y en la cual las búsquedas por similitud sean aproximadas, para lograr reducir la cantidad de accesos a disco y la cantidad de cálculos de distancia.

Resultados y Objetivos

Los resultados obtenidos en los estudios sobre el modelo de espacios métricos además de lograr mejorar el desempeño de los índices métricos analizados, conducen a estudiar su aplicación a otros métodos de acceso [3, 4, 9, 5, 10, 2], lo que se está llevando adelante como distintas tesis de posgrado.

Se espera brindar nuevas herramientas, que administren bases de datos métricas, de manera de acercar a estas bases de datos no convencionales a la

madurez de los modelos tradicionales. Por tal motivo, se continuará el estudio de nuevos diseños de estructuras de datos, que se adapten tanto al nivel de la jerarquía de memorias donde se almacenarán, como a las características de los datos a indexar, con el fin de mejorar su eficiencia en tiempo y espacio. Del mismo modo, se continuará indagando sobre técnicas innovadoras que, sin utilizar índices, permitan resolver consultas eficientemente. Además, se espera mejorar el desempeño de las operaciones de bajo nivel en los DBMS, mediante una nueva arquitectura del procesador.

Actividades de Formación

Dentro de esta línea de investigación se forman alumnos y docentes-investigadores participando en:

- **Maestría en Cs. de la Computación** (UNSL): tesis sobre una versión dinámica eficiente del *DiSAT*.
- **Maestría en Informática** (UNSJ): tesis sobre la evaluación del *knNG* para búsquedas por similitud.
- **Maestría en Informática** (UNSJ): tesis sobre una versión dinámica y para memoria secundaria del *DDiSAT*

Referencias

- [1] E. Chávez, K. Figueroa, and G. Navarro. Effective proximity retrieval by ordering permutations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(9):1647–1658, Sept 2008.
- [2] E. Chávez, M. Di Genaro, N. Reyes, and P. Roggero. Decomposability of disat for index dynamization. *Computer Science & Technology*, pages 110–116, 2017.
- [3] E. Chávez, V. Ludueña, and N. Reyes. Solving all-k-nearest neighbor problem without an index. In *Procs. del XXV Congreso Argentino de Ciencias de la Computación CACIC 2019*, pages 567–576. UniRío editora, 2019.
- [4] E. Chávez, V. Ludueña, N. Reyes, and F. Kasián. All near neighbor graph without searching. *Computer Science & Technology*, 18:61–67, 2018.
- [5] E. Chávez, V. Ludueña, N. Reyes, and P. Roggero. Faster proximity searching with the distal {SAT}. *Information Systems*, 59:15 – 47, 2016.
- [6] E. Chávez, G. Navarro, R. Baeza-Yates, and J. Marroquín. Searching in metric spaces. *ACM Computing Surveys*, 33(3):273–321, September 2001.
- [7] P. Ciaccia and M. Patella. Approximate and probabilistic methods. *SIGSPATIAL Special*, 2(2):16–19, 2010.
- [8] K. Figueroa, C. Martínez, R. Paredes, N. Reyes, and P. Roggero. Dynamic list of clustered permutations on disk. In *Computer Science and Technology Series: XXI Argentine Congress of Computer Science Selected Papers*, pages 201–211. EDULP, 2016.
- [9] K. Figueroa, N. Reyes, A. Camarena-Ibarrola, and L. Valero-Elizondo. Improving the list of clustered permutation on metric spaces for similarity searching on secondary memory. In *10th Mexican Conference on Pattern Recognition (MCPR2018)*, volume 10880, pages 82–92, 2018.
- [10] G. Navarro and N. Reyes. New dynamic metric indices for secondary memory. *Information Systems*, 59:48 – 78, 2016.
- [11] R. Paredes, E. Chávez, K. Figueroa, and G. Navarro. Practical construction of *k*-nearest neighbor graphs in metric spaces. In *Proc. 5th Workshop on Efficient and Experimental Algorithms (WEA)*, LNCS 4007, pages 85–97, 2006.
- [12] H. Samet. *Foundations of Multidimensional and Metric Data Structures (The Morgan Kaufmann Series in Computer Graphics and Geometric Modeling)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2006.
- [13] P. Zezula, G. Amato, V. Dohnal, and M. Batko. *Similarity Search: The Metric Space Approach (Advances in Database Systems)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005. XVIII, 220 p., Hardcover ISBN: 0-387-29146-6.

SERVIDOR PACS USANDO EL ESTÁNDAR DICOM

Repositorio Open Source Web DICOM en Hospital Eleazar Herrera Motta

Lic. Ariel Quiroga Marin¹, Ing. Enrique N. Martinez¹, Ing. Elvio Sigampa Paez¹, Ing. Gabriel, Martinez¹, Cristian Gabriel Palacios¹, Franco David Cardozo¹, Jonatan Maximiliano Padilla¹
Dr. Rivero Jose², Lic. Aguilar Norma², Dr. Carlos Rossi², Ruben H. Quiroga²

¹Departamento de Básicas y Aplicadas, Universidad Nacional de Chilecito 9 de Julio 22, Chilecito, La Rioja, Argentina {cquiroga, emartinez, esigampa, gmartinez}@undec.edu.ar , {cristianpalacios95, franco.undec12, padilla.joni}@gmail.com

²Hospital Eleazar Herrera Motta, Caveno 500, Chilecito, La Rioja
{riverotattoo, aguilarnorma, carlosrossi, rubenhquiroga}@gmail.com

Contexto

La investigación del estándar DICOM nos permitirá conocer, administrar, comprender y adquirir experiencia con un sistema PACS open source. DICOM es quizás el área individual más importante que debe investigar en un sistema de PACS, y este dominio aumenta en dificultad a medida que los estándares DICOM crecen en número y complejidad. En la elección de una solución PACS de código abierto, se deben estudiar varios factores, como el desarrollo, arquitectura y la implementación multiplataforma, para garantizar cumplimiento de los estándares DICOM actuales y futuros.

Este trabajo consiste en la investigación y posible implementación de un Servidor PACS server usando el Estándar DICOM, generando un repositorio Open Source en Hospital Eleazar Herrera Motta de Chilecito, cuyo objetivo es garantizar que las imágenes que generan los distintos equipos de la institución sean de calidad, integrales, oportunos y confiables, lo que comprende la interoperabilidad con otras instancias, así como el desarrollo de indicadores para el monitoreo y la evaluación de la situación de salud del paciente. Aplicar las Tecnologías en el

ámbito de la salud tienen el potencial de mejorar los procesos, optimizar el gasto y contribuir a la seguridad y la equidad en la atención de los pacientes ^(1,2). Se pretende lograr los beneficios potenciales de la interoperabilidad sobre la calidad en la atención médica. Esta situación está ligada a la dificultad en la comunicación y manejo de la información médica, la cual es generada en grandes volúmenes y en diferentes formatos, siendo una problemática a resolver dentro de la institución. De acuerdo al convenio Marco, el Hospital Regional “Eleazar Herrera Motta” de Chilecito que abarca los departamentos de Chilecito, Famatina, Los Sauces, Villa Union, Villa Castelli y Vinchina, proveerá financiamiento al Proyecto, tanto para la infraestructura de Red y la solución de Hardware necesaria.

Líneas de I+D+I

- Computación gráfica, Imágenes y Visualización
- Arquitectura
- Redes y Sistemas Operativos
- Base de Datos
- Cloud Computing
- Minería de Datos

Resumen

DICOM es un estándar ampliamente adoptado para imágenes médicas, pero no es una solución integral para problemas de integración y transacciones en todos los componentes de atención médica, para ello se necesita un sistema para captura y conversión y almacenamiento a formato DICOM de las imágenes de Diagnóstico Médico de los distintos estudios que se realizan en el nosocomio (tomografías, rayos x, ecografías, endoscopias etc.). El acceso al sistema será realizado mediante la intranet o internet, es decir, que se pueda realizar desde cualquier computador conectado en el hospital a la red interna o a su vez conectado a través de la red internet desde cualquier parte del mundo, ya que el PACS funciona en entorno Web DICOM⁽³⁾. que permitirá en un futuro implementar un servidor PACS en la nube. Las imágenes generadas por los distintos equipos, podrían quedar almacenadas en un servidor PACS de la solución open source [arc-light](https://github.com/dcm4che/dcm4chee-arc-light)(<https://github.com/dcm4che/dcm4chee-arc-light>) y además, dichas imágenes quedan totalmente vinculadas a la historia de Salud del Paciente, se acceden netamente a través de DICOM Web (RestFul)⁽³⁾ y visualizadas a través de OHIF Viewer [Oviyam](https://github.com/OHIF/Viewers) (<https://github.com/OHIF/Viewers>), cumpliendo con los estándares DICOM server, HL7 server, Web-based interface.

PALABRAS CLAVE: Integrating HealthCare Enterprise (IHE), Digital Imaging and Communications in Medicine (DICOM), gestión de imágenes digitales, administración de PACS, conformidad con PACS DICOM, Health Level7 (HL7), DCM4CHEE, Docker, arc-light, DICOMWeb (RestFul) y OHIF Viewer Oviyam, Cloud Server, Open-source PACS clients DICOM workstations, OHIF (Open

Health Imaging Foundation) DICOM server Dcm4Che , open source DICOM library.

Introducción

El sistema PACS se integra perfectamente en el entorno hospitalario, a través del estándar de comunicación HL7, contiene los servicios DICOM y HL7 que se requieren para proporcionar almacenamiento, recuperación y flujo de trabajo a un entorno libre.

El PACS DCM4CHE⁽⁴⁾ de código abierto con licencia LGPL permite la rápida adopción de imágenes de un archivo DICOM, estudiar e investigar su arquitectura y las características abarcan los estándares [DICOM, Health Level 7 (HL7), etc.], además de comprender los servicios de DCM4CHEE listados en la tabla (a), incluyen un registro integral de transacciones DICOM, diseñado para dos bases de datos de código abierto y fácilmente disponibles, MySQL y POSTGRES.

Formación de Recursos Humanos

El equipo de trabajo está formado por 4 docentes de las carreras Ingeniería en Sistemas, Licenciatura en Sistemas e Ing. Mecatrónica de la UNDeC (acreditadas por CONEAU), dos de los cuales están realizando de la especialización de Big Data y Postgrado de Informática Medica. También participan profesionales del Hospital Eleazar Herrera Motta y alumnos avanzados de grado, tres de ellos realizando su trabajo de tesina final en esta línea de I+D. Los integrantes son docentes de las asignaturas Sistemas Operativos I y II, Modelos y Simulación, Inteligencia Artificial e informática, Estas asignaturas contemplan la aprobación mediante la

participación en proyectos de investigación, por lo que pueden surgir nuevos trabajos en esta línea.

RESULTADOS ESPERADOS

Para que el Hospital de Chilecito pueda llevar a cabo la gestión y almacenamiento de imágenes médicas de manera eficiente, es necesario que los estudios cumplan con los estándares internacionales que garanticen su reproducción y comunicación como son DICOM y HL7, Digital Imaging and Communications in Medicine y Health Level 7 respectivamente (5), con los que trabajan los sistemas PACS, estos formatos alimentan a los sistemas PACS para que finalmente haya una gestión y resguardo de estos estudios.

Servicios.

Servicio	Descripción
Web-based UI	DCM4CHEE contiene una interfaz de usuario robusta para los administradores que se ejecutan por completo en un navegador Web.
DICOM Storage	Actuando como un archivo, DCM4CHEE es capaz de almacenar cualquier tipo de objeto DICOM a los sistemas de archivo estándar, con la compresión si es necesario.
DICOM Query/Retrieve	Consulta el archivo de objetos DICOM, y los recupera.
WADO and RID	Acceso Web al contenido archivado.
Otros Servicios DICOM	MPPS, GPWL, MWL, Almacenamiento Compromiso, Instancia Disponibilidad de notificación, Estudio Notificación de contenido, contenido de salida para un soporte CD, Protocolos de clasificación, y más.
HL7 Server	Un servidor HL7 integrada que puede actuar sobre ADT, ORM, y los tipos de mensajes ORU.
IHE Services	DCM4CHEE felizmente puede existir en un entorno IHE-capaz mediante la integración de un registro XDS / XDSI y repositorio, que actúa como un nodo de seguro, y proporcionar auditorías conforme.

a) Servicios de DCM4CHEE

El estudio de estos estándares de imágenes digitales de diversos campos, como la radiología, la cardiología, tomografía, endoscopia, ecografía, etc., se pueden combinar armoniosamente en un PACS server con la información de los sistemas de información clínica y otras bases de datos. Otra ventaja de utilizar esta solución PACS de código abierto es que se integra con equipos de proveedores líderes en equipamiento de imágenes DICOM y proporcionar acceso en sus estaciones de

trabajo en toda la institución.

La investigación del estándar DICOM y PACS Server nos permitirá brindar una solución Open Source de bajo costo y resuelve el problema que tiene el Hospital de Chilecito, interconectando todos los servicios esenciales permitiendo un enlace práctico, dinámico y almacenamiento seguro, permitiendo el flujo de información y brindar diagnósticos más precisos en menos tiempo.

Referencias

- [1] Institute of Medicine (U.S.). Committee on Quality of Health Care in America. Crossing the quality chasm : a new health system for the 21st century. Washington, D.C.: National Academy Press; 2001.
- [2] Institute of Medicine (U.S.). Committee on Patient Safety and Health Information Technology. Health IT and patient safety: building safer systems for better care. Washington, DC: National Academies Press; 2012.
- [3] DICOM® Publications and DICOMWeb™ Publications are published by and copyright owned by the National Electrical Manufacturers Association <https://www.dicomstandard.org/current>. The DICOM Standard. "DICOM Homepage." Retrieved Sep 2, 2016, from <http://dicom.nema.org/standard.html>.
- [4] Open Source Clinical Image and Object Management, link: <https://www.dcm4che.org/>
- [5] Organización Mundial de la Salud. eHealth standardization and interoperability. 132ª Asamblea mundial de la Salud. (resolución EB132.R8). Ginebra (Suiza), 2013 [cited 2013 30 Sep]; Available from: <http://goo.gl/05qyP1>

DETECCIÓN DE YEMAS BROTADAS PARA LA ESTIMACIÓN TEMPRANA DEL RENDIMIENTO DE UNA PLANTACIÓN DE KIWI

Silvia N. Pérez¹, Gustavo A. Dejean¹, Mónica Giuliano¹, Alejandra Yommi³, María A. David³, Natalia L. Murillo⁴, Federico Balaguer², Ignacio A. García Ravlic¹, Dante H. Mendoza¹

¹Escuela de Informática, Universidad Nacional del Oeste

²Stream S.A., ³INTA, Estación Experimental Balcarce

⁴INTA, Agencia de Extensión Rural Otamendi

{sperez; gdejean; mgiuliano; igargia; hmendoza} @uno.edu.ar;

{yommi.alejandra; david.maria; murillo.natalia} @inta.gob.ar; federico.balaguer@digital-stream.com.ar

RESUMEN

Se presentan aquí resultados de detección de órganos vegetales en imágenes usando redes convolucionales, correspondientes a la primera etapa de un proyecto que propone desarrollar procedimientos para estimar en forma temprana el rendimiento de un lote de producción de kiwi. Para lograr la predicción del rendimiento, se consideran tres estadios fenológicos del crecimiento del kiwi y para validar, se dispondrá de resultados finales de cosecha. Para cada etapa se consideran 1000 imágenes, capturadas en toda el área cubierta por el cultivo, que permiten entrenar y elegir un modelo para conteo automático de objetos.

Además, el modelo se evalúa en 50 sitios elegidos al azar en la plantación, donde se compara el conteo manual con el conteo automático obtenido sobre imágenes de cada sitio. Los resultados en esta primera etapa, realizada sobre yemas brotadas en 47 sitios de la plantación, son promisorios en cuanto a la utilización de modelos de deep learning para detección de objetos y muestran la necesidad de profundizar en el ajuste de modelos y estrategias de conteo. Esto permite inferir que puede utilizarse el conteo automático como input en el modelo predictivo final. La metodología propuesta permitirá la predicción del rendimiento total de la plantación.

Palabras clave: *Procesamiento de Imágenes, Detección de objetos, Deep Learning, Estimación de producción.*

CONTEXTO

La línea de investigación que se presenta es parte del proyecto de investigación “Estimación temprana de rendimiento del cultivo de kiwi mediante el procesamiento de imágenes” que se desarrolla en forma conjunta por la Universidad Nacional del Oeste (UNO), el Instituto Nacional de Tecnología Agropecuaria (INTA) y la empresa Stream S.A. Cada una de estas entidades aborda los aspectos relacionados con su área de especialización [Dejean et al, 2021].

1. INTRODUCCIÓN

El kiwi (*Actinidia chinensis* var. *deliciosa*) es una planta introducida en Argentina en los años 80's, con una superficie en producción que se encuentra en crecimiento sostenido desde el año 2004. Actualmente se estima que hay cerca de 1000 ha plantadas de kiwi en el país, de las cuales 600 ha se sitúan en el sudeste de la provincia de Buenos Aires. Esta zona es la principal área productiva de kiwi, aportando más de la mitad del volumen nacional. [David et al, 2018, 2020].

El fruto de kiwi se posiciona internacionalmente como una fruta con muchos beneficios para la salud, lo que logró instalarlo en todo el mundo. En nuestro país se estima un consumo de medio kilo por habitante y por año [Ministerio de Agricultura, Ganadería y Pesca, 2020]. La producción se destina principalmente al mercado interno, aunque cubre sólo el 50% de la demanda total y el resto a la exportación (3000 tn en el año 2020). La importancia creciente del cultivo motiva investigaciones, fundamentalmente del INTA, que permitan

lograr un producto de calidad según estándares internacionales, a la vez que incrementar la producción local.

El éxito de una plantación de kiwi depende en gran medida de la cantidad de frutos cosechados por hectárea y de la proporción obtenida de frutos medianos a grandes, debido a que estos tamaños suelen tener precios diferenciales, mejorando el retorno económico [Testolin y Costa, 1992]. La cantidad de frutos obtenidos se sabe dependiente del número de yemas, del porcentaje de brotación de las yemas, del porcentaje de yemas brotadas que son reproductivas y del número de inflorescencias por brote reproductivo [Buwalda y Smith, 1988]. Esto motiva considerar como variables de estudio para la predicción del número de frutos, al conteo obtenido en cada una de tres estadios fenológicos: yemas brotadas (brotes/hojas), botones florales y frutos pequeños. El conteo automático de estos órganos frutales permite aprovechar la tecnología para lograr resultados más rápidos y eficientes en la toma de datos.

La utilización de tecnologías de detección de objetos en imágenes es una tendencia creciente en el área de producción frutihortícola [Sa 2016; Xia, 2022; Hussain, 2022]. Estos utilizan redes neuronales convolucionales profundas (DCNN, por sus siglas en inglés *Deep Convolutional Neural Networks*) para estimación de rendimiento de cultivos frutales, entrenando modelos para reconocimiento de objetos (frutos) en imágenes. Particularmente en producción de kiwi, varias publicaciones dan cuenta de la detección automática de frutos en una plantación como un desafío para la recolección por medios mecánicos automatizados [Williams et al, 2019; Song et al, 2019]. Una aplicación importante de las técnicas de DCNN corresponde a la detección de estadios tempranos del fruto con el objetivo de predecir la producción [Lim et al, 2020]. Actualmente, el método más utilizado para estimar el volumen de cosecha es el conteo visual de los frutos por unidad de superficie en una etapa avanzada. Esta modalidad conlleva dificultades en grandes plantaciones, así como también afecta la previsibilidad operativa dada

la cercanía con la etapa de cosecha. Una estimación temprana de la producción a obtener permitiría ajustar de modo sustentable los insumos necesarios para la producción, gestionar anticipadamente las necesidades operativas para la cosecha y controlar de modo más eficiente la capacidad de almacenamiento, así como gestionar la venta y distribución.

El brote de las yemas y el desarrollo de los botones florales y frutos a lo largo de la temporada es influenciado por las horas de frío acumuladas durante el periodo invernal, el perfil de heladas (duración e intensidad) y la cantidad de horas de stress térmico (considerando temperatura y humedad). Estas variables climáticas deben considerarse al momento de crear un modelo de predicción temprana del rendimiento. La obtención de este modelo a partir de los estadios más tempranos (yemas brotadas y botones florales) tendría mayores ventajas para el productor, aunque probablemente sea más complejo dado la mayor influencia de las variables climáticas.

En este proyecto se propone la identificación, conteo y seguimiento de brotes, botones florales y frutos pequeños de una plantación de kiwi, mediante la utilización de métodos de deep learning para detección de objetos en imágenes. El objetivo final de esta tarea es obtener estimaciones tempranas de rendimiento de producción de kiwi. Para esto, trabajan de modo colaborativo profesionales de la Fruticultura juntamente con otros de Informática y Procesamiento digital de la información.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La línea de investigación aquí presentada da una nueva perspectiva para UNO, correspondiente al procesamiento de imágenes con redes convolucionales (DCNN).

El proyecto propone estimar la cantidad de frutos de kiwi a producir en base al procesamiento de imágenes para detección y conteo automático de objetos en la plantación. Se suman a estas, mediciones obtenidas por sensores con datos de temperatura del aire,

humedad relativa, entre otras mediciones en la plantación. La detección de objetos en imágenes se realiza en cada una de tres etapas fenológicas asociadas a estados tempranos de desarrollo del cultivo: brotes/hojas, botones florales y frutos pequeños. Para esto se emplean modelos predictivos de *deep learning* para la detección de objetos, generando el conteo automático según cada etapa, los que se comparan con el conteo manual in situ.

La plantación de kiwi en estudio ocupa 2,2 hectáreas y se ubica cercana a la ciudad de Miramar, en la Provincia de Buenos Aires. Para lograr los objetivos del proyecto, se tomaron imágenes en distintos estadios fenológicos del cultivo, recorriendo toda la superficie de la plantación. Además, se colocaron al azar 50 marcos de 1 metro de ancho por 2,5 metros de largo, que se fijaron en plantas femeninas, donde se realizan los conteos manuales e imágenes. Los marcos se distribuyen según criterios de muestreo de la producción de la plantación.

Para cada una de las tres etapas de desarrollo del kiwi se consideraron dos tipos de imágenes: las de tipo A se corresponden a imágenes en variados puntos de la plantación en diferentes momentos del día con condiciones de iluminación variada. Las imágenes tipo B se corresponden con los 50 marcos y son utilizadas para comparación del modelo automático con el conteo manual. Las imágenes tipo A son aproximadamente 1000 por cada etapa y las de tipo B son tres por cada marco. El primer conjunto de imágenes es etiquetado manualmente para luego entrenar el modelo y aplicarlo a la detección de objetos en las imágenes de tipo B. Con esto, las imágenes tipo B permiten evaluar el modelo obtenido en contraste con la “verdad de campo” (conteo manual).

Las imágenes tipo A son etiquetadas por un operador y luego se separan al azar en tres conjuntos: entrenamiento, validación y prueba (65%, 34% y 1% respectivamente). Esto significa que cada una de esas imágenes tiene un cuadro delimitador para cada objeto encontrado por el operador.

En cada una de las tres etapas fenológicas, de forma independiente, se consideran modelos de redes convolucionales pre-entrenados, los que luego se ajustan a los datos propios utilizando los conjuntos de entrenamiento y validación, recursivamente hasta lograr convergencia en parámetros de pérdida. La evaluación de la calidad de cada uno de estos modelos se realiza analizando la detección sobre imágenes de prueba, las que se comparan con las imágenes originales. Esto permite seleccionar un modelo entre todos los modelos en competencia utilizando métricas usuales de calidad. Posteriormente, se comparan los conteos de detecciones obtenidos por el modelo elegido sobre las imágenes tipo B, con el conteo manual realizado por otros operadores en la misma plantación de kiwi.

En el proceso de detección se fija como probabilidad mínima para considerar detectado un objeto, un valor de corte de 0,5.

Para evaluar la calidad de un modelo se consideran tres métricas básicas: precisión y sensibilidad o *recall*, las cuales se definen como

$$\text{Precisión} = \text{VP} / (\text{VP} + \text{FP}) \quad \text{Recall} = \text{VP} / (\text{VP} + \text{FN})$$

donde:

VP: verdaderos positivos (yemas con brotes/hojas reales que fueron bien predichos por el modelo);

FP: falsos positivos (predichos erróneamente como yemas con brotes/hojas, también llamados detecciones falsas)

FN: falsos negativos (yemas con brotes/hojas que no fueron detectados por el modelo, también llamadas detecciones pérdidas)

Otra métrica de interés para medir la calidad del modelo es F1, que combina precisión y recall y se define como:

$$F1 = 2 * \text{precisión} * \text{recall} / (\text{precision} + \text{recall})$$

El procedimiento descrito permite elegir un modelo para el conteo automático en cada una de las tres etapas fenológicas. Estos conteos serán utilizados como input del modelo predictivo final. El conteo de frutos de la cosecha de cada uno de los marcos será la variable respuesta del conjunto de datos. Los resultados a obtener permitirán la predicción del rendimiento total de la plantación. Se

podrán jerarquizar los modelos según etapa fenológica y además mejorarlos sumando otras variables, como por ejemplo meteorológicas.

3. RESULTADOS OBTENIDOS Y ESPERADOS

Hasta el momento se completó la obtención de imágenes A y B, además de conteos manuales, para las tres etapas fenológicas (brotes/hojas, botones florales y frutos pequeños).

Con los conjuntos de entrenamiento y validación de imágenes tipo A de brotes/hojas, se entrenaron diferentes modelos que fueron validados apropiadamente. Para la elección del modelo a utilizar en el conteo automático en las imágenes B, se tomó en cuenta la métrica F1 la cual combina precisión y *recall*. El mejor valor obtenido fue $F1=0,915$.

La Fig. 1 muestra la detección de yemas con brotes/hojas realizada por el modelo seleccionado en una porción de uno de los marcos. La Fig. 2 muestra un detalle de la detección de un brote, donde se aprecia la probabilidad con la que fue detectado como tal.

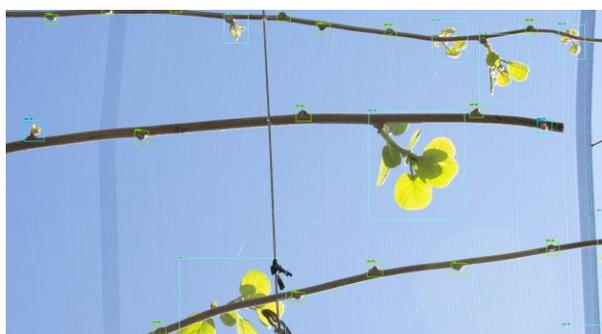


Fig. 1: Yemas con brotes/hojas detectadas por el modelo.



Fig.2. Detalle de la Fig.1 donde se observa la probabilidad con que se detectó el brote.

El conteo automático realizado sobre cada marco (Mo) se comparó con el conteo manual in situ (Ma). Las diferencias entre ambos valores, relativas al conteo manual, permite evaluar el error del modelo en cada imagen. De los 50 marcos de la plantación se juzgaron útiles sólo 47. En la Tabla 1 se observa los valores promedio de conteo automático y conteo manual, 23,60 y 26,68, respectivamente, así como el promedio de la diferencia relativa resultó de -0,10.

Medida	conteo del modelo (Mo)	conteo manual (Ma)	$\frac{(Mo-Ma)}{Ma}$
Promedio	23,60	26,68	-0,10
Mínimo	9	13	-0,44
Máximo	37	44	0,75

Tabla 1: Comparación relativa de conteos (n=47)

En la etapa de yemas con brotes/hojas se observa que el modelo automático subestima el conteo en general en los marcos, sin embargo, en el 83% de los mismos se obtuvo un error en valor absoluto inferior a 0,25. Dado que se observaron algunas diferencias importantes entre ambos conteos, se realizó una revisión de los casos particulares lo que sugiere revisar las estrategias del conteo manual. Esto motiva la necesidad de replicar el ensayo en la próxima temporada introduciendo un control mayor en los conteos en los marcos.

La acumulación de horas de frío invernales y las temperaturas en primavera inciden sobre la brotación y la diferenciación de las yemas. En años en que se acumulan al menos 950 horas de frío y las temperaturas son cálidas, se espera una alta brotación de yemas, con predominancia de yemas reproductivas que darán fruta, respecto a las vegetativas, que sólo producen brotes y hojas. Se evaluará la influencia de las horas de frío acumuladas y de las temperaturas, así como la ocurrencia de heladas, en las predicciones de rendimiento basadas en el conteo automático de yemas para considerarlas en el modelo predictivo.

Se continuará con la evaluación de modelos predictivos a partir de la detección automática y manual de botones florales y de frutos

pequeños. Considerando alguno de los estadíos fenológicos en estudio, se espera lograr un modelo predictivo para estimar tempranamente el rendimiento (n° de frutos/ha) de la producción de kiwi a partir de imágenes, que a su vez podrá mejorarse en las siguientes temporadas.

En la etapa cumplida hasta el momento se obtuvieron resultados promisorios al aplicar algoritmos de *deep learning* para la detección de yemas brotadas.

La estimación temprana del rendimiento de un cultivo de kiwi fue planteada como una necesidad por productores del sudeste de la provincia de Buenos Aires. Los productores, a través del INTA, tendrán una herramienta para estimar la producción y contribuir con una mejor gestión de los recursos ambientales y de los insumos. Esto permitirá mejorar la gestión de la cosecha, el almacenaje de la fruta y su comercialización, y también beneficiará a empacadores y exportadores.

Se considera que, más allá de cumplidos los objetivos de estimación temprana del rendimiento, esta línea de investigación permite formar un equipo de trabajo capaz de proponer nuevos proyectos relacionados con detección automática de objetos, aplicables en otros cultivos frutihortícolas.

4. FORMACIÓN DE RECURSOS HUMANOS

En el proyecto participan siete investigadores formados y dos becarios alumnos. Estos últimos se inician de este modo en la investigación en el área de informática aplicada. El trabajo realizado ha permitido también la realización del trabajo final de carrera de la Licenciatura en Informática de uno de los alumnos becarios.

5. BIBLIOGRAFÍA

[Buwalda y Smith, 1988] Buwalda, J.G. and Smith, G.S. 1988. A mathematical model for predicting annual fertiliser requirements of kiwifruit vines. *Scientia Hort.* 37:71-86.

[David et al, 2018] David, M.A., Yommi, A., Sánchez, E. 2018. Fenología del cultivo de kiwi en el sudeste de Buenos Aires.

<https://inta.gob.ar/documentos/fenologia-del-cultivo-de-kiwi-en-el-sudeste-de-buenos-aires>.

- [David et al, 2020] David, M.A., Yommi, A., Sanchez, E. 2020. Elección del terreno y plantación del cultivo de kiwi. 1a ed. Balcarce, Buenos Aires: Ediciones INTA. Libro digital, 39 p.
- [Dejean et al, 2021] Dejean, G., Balaguer, F., Yommi, A., Doorn, J., David, A., Murillo, N., García I., Mendoza, D. Integración del Procesamiento Imágenes e Internet de las Cosas en la estimación temprana del rendimiento de cultivos frutales. XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021, Univ. Nac. de Chilecito).
- [Hussain et al, 2022] Hussain, D.; Hussain, I.; Ismail, M.; Alabrah, A.; Ullah, S.; Alaghbari, H. (2022) A Simple and Efficient Deep Learning-Based Framework for Automatic Fruit Recognition, *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 6538117, 8 pages, 2022.
- [Lim et al, 2020] Lim, J.; Ahn, H.S.; Nejati, M.; Bell, J.; Williams, H.; MacDonald, B.A. Deep Neural Network Based Real-time Kiwi Fruit Flower Detection in an Orchard Environment. arXiv 2020, arXiv:2006.04343.
- [Ministerio de Agricultura, Ganadería y Pesca, 2020] Ministerio de Agricultura, Ganadería y Pesca, 2020. Cadena de kiwi. 16 págs. http://www.alimentosargentinos.gob.ar/HomeAlimentos/Cadenas%20de%20Valor%20de%20Alimentos%20y%20Bebidas/informes/Resume_n_Cadena_KIWI_ENERO_2020.pdf
- [Sa et al, 2016] Sa I, Ge Z, Dayoub F, Upcroft B, Perez T, McCool C. DeepFruits: A Fruit Detection System Using Deep Neural Networks. *Sensors*. 2016; 16(8):1222.
- [Song et al, 2019] Song, Z., Fu, L., Wu, J., Liu, Z., Li, R., & Cui, Y. (2019). Kiwifruit detection in field images using Faster R-CNN with VGG16. *IFAC-PapersOnLine*, 52(30), 76-81.
- [Testolin y Costa, 1992] Testolin, R., and Costa, G. 1992. Modelling a kiwifruit orchard. *Acta Hort.* 313:99-103.
- [Williams et al, 2020] Williams, H. A., Jones, M. H., Nejati, M., Seabright, M. J., Bell, J., Penhall, N. D., ... & MacDonald, B. A. (2019). Robotic kiwifruit harvesting using machine vision, convolutional neural networks, and robotic arms. *biosystems engineering*, 181, 140-156.
- [Xia et al, 2022] Xia, X.; Chai, X.; Zhang, N.; Zhang, Z.; Sun, Q.; Sun, T. Culling Double Counting in Sequence Images for Fruit Yield Estimation. *Agronomy* 2022, 12, 440.

Sistema de Información basado en Norma Dicom para aplicaciones oftalmológicas orientadas a Retinopatías del Prematuro

Adrián Salvatelli, Alejandro Hadad, Gustavo Bizai, Diego Evin

Laboratorio de Sistemas de Información, Departamento Académico Informática,
Facultad de Ingeniería, Universidad Nacional de Entre Ríos
Ruta 11, Km 10,5, Oro Verde, Entre Ríos, 0343-475100, Interno 149
{adrian.salvatelli, alejandro.hadad, gustavo.bizai, diego.evin}@uner.edu.ar

Resumen

La adquisición de imágenes para la evaluación de la Retinopatía del Prematuro presenta desafíos tanto en lo que refiere a la calidad de la imagen como en la estandarización hacia el estándar Dicom para su conformación como modalidad. En este sentido este proyecto desarrolla e implementa un desarrollo para la estandarización de las imágenes y videos provenientes de los estudios oftalmológicos de pacientes prematuros y aborda la problemática de preprocesamiento y procesamiento de dichos estudios a fin de mejorar la calidad para la mejora de los procesos de toma de decisiones.

Palabras clave: Retinopatía del Prematuro, Adquisición de Imágenes, Estandarización, Preprocesamiento

Contexto

La línea de I+D+i presentada es parte del Proyecto de Investigación y Desarrollo PIDUNER N°6205 “Sistema de Información basado en Norma Dicom para aplicaciones oftalmológicas

orientadas a Retinopatías del Prematuro (ROP)”, financiado por la Universidad Nacional de Entre Ríos y ejecutado en el ámbito de la Facultad de Ingeniería (FI-UNER) [1, 2, 3].

Introducción

La Retinopatía del Prematuro (ROP) es una enfermedad del desarrollo de los vasos retinianos y el vítreo, con anormal maduración y diferenciación celular en una retina que no ha completado su maduración debido a que el bebé nació prematuramente. Esta patología, comparte características fisiopatológicas comunes con otras retinopatías proliferativas asociadas a isquemia focal y aparición consiguiente de neovascularización, cómo la retinopatía diabética [4].

Puede prevenirse, en la mayoría de los casos, en Unidades de Cuidados Intensivos Neonatales (UCIN) con infraestructura adecuada y personal suficiente y capacitado que realice un correcto manejo de conductas perinatales y que controle los factores de riesgo mencionados (prevención primaria). El diagnóstico oftalmológico en el momento adecuado a través de la pesquisa

sistemática y el tratamiento oportuno mejoran el pronóstico visual de los niños (prevención secundaria) [5].

Estas pesquisas deben realizarse con mucha frecuencia, durante el seguimiento del paciente, y posterior a su alta médica. Esto supone un gran número de imágenes e información valiosa que debe registrarse, y realizar un seguimiento de cada paciente.

Factores que afectan la calidad

Diagnóstica

Para la realización del estudio, se utiliza el Oftalmoscopio Binocular Indirecto (vincha de cabecera) OBI, contra sistemas de imágenes de contacto de gran angular (por ejemplo, RetCam; Clarity Medical Systems, Pleasanton, CA, EE. UU). El primero no permite un registro y menos aún una vinculación con registros médicos electrónicos. El segundo es un sistema dedicado con poder de integración fácilmente a una red de telemedicina. Obviamente la calidad en el poder diagnóstico entre estos dos equipos es muy diferente.

La calidad de la imagen es un factor importante para una detección ROP efectiva y confiable. Esto es particularmente un problema en los fondos con pigmento oscuro o en bebés con pequeñas fisuras palpebrales, que impiden el contacto adecuado de la cámara y la superficie de la córnea (RetCam). La neblina vítrea o corneal debida a prematuridad extrema, hemorragia vítrea o artefactos de movimiento también son factores adicionales que pueden afectar la calidad de la imagen, lo que dificulta su interpretación, ocurriendo esto en muchos casos durante el examen clínico también con OBI Las imágenes no interpretables

oscilan entre el 8% y el 21% en los estudios publicados [6,7]. Este proyecto se desarrolla a partir de un dispositivo tipo OBI con una posterior etapa de digitalización [1].

Por otro lado el estándar DICOM (Digital Imaging and Communications in Medicine) es el estándar reconocido mundialmente para el intercambio de imágenes médicas, pensado para su manejo, visualización, almacenamiento, impresión y transmisión. Incluye la definición de un formato de fichero y de un protocolo de comunicación de red. El protocolo de comunicación es un protocolo de aplicación que usa TCP/IP para la comunicación entre sistemas. Los ficheros DICOM pueden intercambiarse entre dos entidades que tengan capacidad de recibir imágenes y datos de pacientes en formato DICOM.

En oftalmología la AAO (American Academy of Ophthalmology) trabaja en él: WG-09 (Dicom Working Group 09): Ophthalmology [8], ha trabajado y documentado en los datos necesarios en las clases de comunicación DICOM. De esta manera asegura que los equipos utilizados en oftalmología puedan ser integrados a un PACS (Picture Archiving and Communications System) físico o virtual.

Líneas de Investigación, Desarrollo e Innovación

La investigación pertenece al Área Computación Gráfica, Imágenes y Visualización. En relación a la temática específica del proyecto las líneas de trabajo están vinculadas a dos ejes. Por un lado la estandarización norma Dicom adecuando la modalidad de toma de imágenes es oftalmología a través de dispositivos móviles. Por el otro el

análisis de diferentes técnicas de Procesamiento Digital de Señales y de Inteligencia Artificial orientado a la mejora de calidad de imágenes y video a partir de las dificultades que plantea la escena de adquisición de las mismas [1, 2, 3].

Resultados y Objetivos

En las primeras etapas del proyecto se trabajó en dos ítems muy relevante para el desarrollo. El primero fue el análisis y búsqueda de tecnologías que pudieran ser aplicadas en la implementación del prototipo de software smartphone bajo estándares. El segundo fue el análisis de los primeros videos e imágenes adquiridas, para luego desarrollar una estrategia de implementación para el procesamiento digital de video e imágenes capturadas.

En este sentido se analizaron los recursos existentes con mayor adopción en la actualidad para aplicaciones ROP y otras de similares características. Con dicha información se decidió implementar la aplicación “Smartphone Fundoscopy” para las siguientes tareas:

- Toma imágenes o videos y los convierte a un archivo DICOM.
- Etiqueta el estudio como modalidad OP (ophtalmology).
- Brinda la posibilidad que el profesional pueda buscar de estudios por paciente, fecha de realización, modalidad.
- Cuenta con un visualizador integrado de imágenes DICOM.
- Flexibiliza la posibilidad de ajuste de la aplicación a las pantallas tanto de un Smartphone como a las de PC

Para ello fue necesario definir la arquitectura de Smartphone Fundoscopy, las tecnologías utilizadas, la base de datos, el Sistema de Comunicación y Archivado de Imágenes PACS (Picture Archiving and Communication System), y el lenguaje de programación para llevar las imágenes al estándar DICOM.

En relación al preprocesamiento de las imágenes se ha separado en fotogramas los videos ODI ROP que incluyen el video en crudo del procedimiento, los fotogramas que a su criterio muestran la patología y una planilla donde se establece el diagnóstico ROP según la clasificación estándar. La mayoría de los videos tienen una resolución digital de 3840 x 2160 (8 Mpíxels aprox.) con una duración de entre 7 y 32 segundos en formato MOV y MP4. Un 10% de estos estudios, fueron realizados con resolución digital de 1920 x 1080 (2 Mpíxels). Ambos estudios fueron obtenidos con un equipo iPhone 8. Iphone y Ipad son dos dispositivos móviles aceptados para uso médico de imágenes por FDA, quien establece una resolución digital mínima, de la imagen completa capturada para uso oftalmológico de 4 Mpixels. La cantidad de fotogramas seleccionados por los expertos varía entre 3 y 4 mayormente representativos por video capturado. El formato del fotograma es RGB con 8 bits por canal de color.

Dentro de estos fotogramas, solo una parte son píxeles de interés para un futuro diagnóstico. Estos “píxeles útiles” se encuentran dentro de la imagen que proyecta la lupa de 28 o 66 Dioptrías. Para un análisis cuantitativo de esta cantidad de píxeles útiles se ha desarrollado scripts que segmentan dicha lupa de cada fotograma y enmascara con color negro la información irrelevante.

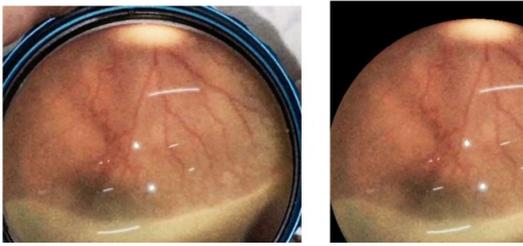


Figura 1 - Imagen con ROP original (izq.) y procesada (der.)

Analizando cada fotograma procesado y recortado se ha determinado que el 58% de los fotogramas tienen una resolución digital de 2 Mpixels, mientras que el resto una resolución muy inferior. De este 58% la cantidad de píxeles útiles es inferior a 1,2 Mpixels, es decir un promedio aproximado del total de píxeles del %55.

En este mismo análisis se han incluido parámetros obtenidos de los histogramas de los píxeles útiles, resultando sus medias entre los colores 60 y 140 con desvío estándar promedio de 77 valores.

Morfológicamente los histogramas presentan más de 3 modas y los rendimientos de color de cada fotograma es muy disperso.

En una etapa posterior se trabajó sobre distintos modelos gaussianos de PSF para ser utilizados como filtros de deconvolución. De esta manera se han quitado algunas aberraciones ópticas mejorando la calidad visual de la imagen.

Actualmente se está trabajando en una secuencia o pipeline de procesamiento, que consta de etapas de enmascarado de imágenes, registro, normalización, remoción de artefactos, submuestreo, realce de vasos y ensamblado, para la conformación de un mosaico de las imágenes de la retina. De estas tareas, comenzó con la etapa de normalización: todas las imágenes y videos se llevan a un

formato estándar y se asegura de que los niveles de intensidad de las imágenes se encuentren dentro de los marcos definidos.

Luego, se centró en la clasificación de los cuadros informativos (que contienen imagen del fondo ocular) para desechar los no informativos (sin información diagnóstica, lo que sucede durante la manipulación de los párpados del bebé y búsqueda del cuadro de imagen). Comenzó haciendo un análisis en el espacio de colores HSV, referido a tonalidad (Hue), saturación (Saturation) e intensidad (Value). La clasificación en el espacio HSV es más robusta a los cambios de iluminación, sombras y las variaciones de textura que en otros espacios de color. Se especifica un contorno cerrado S en el espacio HSV, donde los píxeles con valores dentro del contorno, se clasifican como retinianos, y se obtiene un puntaje. La puntuación HSV de un fotograma viene dada por la proporción de píxeles de la retina en la imagen con respecto a la cantidad de píxeles totales de la imagen.



Figura 2 - Imagen original a la izquierda. Imagen con máscara HSV

Completadas las actividades de preprocesamiento orientadas a mejorar la calidad de imagen se procederá a las actividades de procesamiento de las mismas con el fin de detectar las regiones patológicas en imágenes provenientes de pacientes con ROP.

Formación de Recursos Humanos

El equipo de trabajo está conformado por especialistas del área de sistemas de información, inteligencia artificial y bioingeniería. Relacionadas con el Proyecto de Investigación se están generando distintas actividades a nivel de Grado y Posgrado.

Referencias

- [1] “Análisis de un Sistema de Información para Retinopatías del Prematuro (ROP)”, A.Salvatelli, A. Hadad, D. Evin, G. Bizai, B. Franseschini, B. Drozdowicz, *Revista Argentina de Bioingeniería*, VOL. 24 (3), 2020, pag. 25 – 30. <http://revistasabi.fi.mdp.edu.ar/index.php/revista/article/view/317/354>, Published: 2020-11-09. ISSN 2591-376X.
- [2] Torres, R. M., Saidman, G., Monteoliva, G., & Salvatelli, A. (2020). Networking by WhatsApp: smartphones and their use for Retinopathy of Prematurity (ROP) in Argentina and Latin America. *Investigative Ophthalmology & Visual Science*, 61(9), PB00112-PB00112.
- [3] Torres, R. M., Monteoliva, G. A., Saidman, G., Salvatelli, A., Alazard, G., Ortiz-Basso, T., ... & Ghersinich, A. (2019). A novel tool for the retinopathy of prematurity evaluation, associated with a digital indirect ophthalmoscope hands-free device (DIO-HF). *Investigative Ophthalmology & Visual Science*, 60(11), PB032-PB032.
- [4] Retinopatía del Prematuro, Ana Campo Gesto y col, editorial Marban, 2011, cap. 1 , pag. 73, ISBN: 9788471017154.
- [5]. Retinopatía de prematuro en Argentina. Dra. Celia Lomuto, Médica Pediatra Neonatóloga; Coordinadora Grupo ROP Argentina y Programa Nacional de Prevención de la ceguera en la Infancia por ROP. <https://vision2020la.wordpress.com/2015/04/01/retinopatia-del-prematuro-en-argentina/>
- [6] Impact of number and quality of retinal images in a telemedicine screening program for ROP: results from the e-ROP study, David Morrison, MD,^a Erick D. Bothun, MD,^b Gui-Shuang Ying, PhD,^c Ebenezer Daniel, MBBS, MS, PhD,^c Agnieszka Baumrijer, MS,^d and Graham Quinn, MD, MSCE,^{c,d} for the e-ROP Cooperative Group, *Journal of AAPOS*, <http://dx.doi.org/10.1016/j.jaapos.2016.08.004>, Oct. 2016
- [7] Screening and treatments using telemedicine in retinopathy of prematurity (Review), Aristomenis Thanos, Yoshihiro Yonekawa, Bozho Todorich, Darius M Moshfeghi, Michael T Trese, *Eye and Brain Dove press Journal*, 17 August 2016, vol. 8 147–151, available at <https://www.dovepress.com/eye-and-brain-journal>, DOI <https://doi.org/10.2147/EB.S94440>
- [8] WG-09 (Dicom Working Group 09): Ophthalmology, publicación Online, <https://www.dicomstandard.org/wgs/wg-09/>

EL APOORTE DE LAS REALIDADES ALTERNATIVAS AL OBJETIVO DE DESARROLLO SOSTENIBLE DE SALUD Y BIENESTAR

G. Rodríguez, N. Jofré, Y. Alvarado, J. Fernandez, R. Guerrero
Laboratorio de Computación Gráfica / Dpto. Informática / FCFMyN
Universidad Nacional de San Luis
Ejército de los Andes 950, Tel: 0266 4420823, San Luis, Argentina
{gbrodriguez, npasinetti, ymalvarado, jmfer, rag}@unsl.edu.ar

RESUMEN

El uso de las Realidades Alternativas se ha extendido a numerosas actividades desde la creación de las mismas. No obstante, en los últimos tiempos y como consecuencia de la maduración de las tecnologías asociadas y la pandemia COVID-19, se han convertido en una fuente de riqueza no sólo en el propio desarrollo de herramientas sino también en la conformación de acciones futuras asociadas a mejorar las vidas y las perspectivas de las personas en todo el mundo.

La Realidad Virtual y la Realidad Aumentada son dos de las tecnologías más emergentes y disruptivas existentes al día de hoy. Los costes para su uso se han abaratado significativamente a lo largo de los últimos años y su uso es cada vez mayor en áreas como psicología clínica, neurociencias, psicoterapia y salud mental, dado que permiten inducir emociones, reacciones y pensamientos que son prácticamente idénticos a los que se darían en una situación real. Según las Naciones Unidas, para lograr los Objetivos de Desarrollo Sostenible, es fundamental garantizar una vida saludable y promover el bienestar universal. En la presente propuesta de trabajo se propone identificar tres grandes líneas en las cuales las Realidades Alternativas pueden dar soporte al cumplimiento de dicho objetivo.

Palabras clave: Objetivos de Desarrollo Sostenibles, Realidad Virtual, Realidad Aumentada, Interacción Humano Computadora, Empatía computacional.

CONTEXTO

La propuesta de trabajo se lleva a cabo dentro del proyecto “*Realidades Alternativas como lenguaje generativo aplicado a la solución de problemas reales*”. Este proyecto es desarrollado en el ámbito del Laboratorio de Computación Gráfica de la Universidad Nacional de San Luis.

1. INTRODUCCIÓN

Garantizar una vida sana y promover el bienestar para todos en todas las edades es importante para la construcción de sociedades prósperas. Sin embargo, a pesar de los importantes avances realizados en los últimos años en cuanto a la mejora de la salud y el bienestar de las personas, todavía persisten desigualdades en el acceso a la asistencia sanitaria. La mitad de todas las personas de las regiones en desarrollo tienen acceso a la asistencia sanitaria que necesitan, lo que provoca que cada año mueran más de 6 millones de niños antes de los 5 años. Epidemias como el VIH/SIDA medran, donde el miedo y la discriminación limitan la capacidad de las personas para recibir los servicios que necesitan a fin de llevar una vida sana y productiva. El acceso a la salud y el bienestar es un derecho humano, y esta es la razón por la que la Agenda de la ONU para el Desarrollo Sostenible [1] ofrece una nueva oportunidad de garantizar que todas las personas, no solo las de mayor poder adquisitivo, puedan acceder a los más altos niveles de salud y asistencia sanitaria.

Las realidades alternativas provistas por la Realidad Virtual (RV) y la Realidad Aumentada (RA) poseen usos que pueden ser muy útiles para el cumplimiento de los Objetivos de Desarrollo Sostenible (ODS) establecidos en la agenda, con un coste marginal mínimo, por lo que se convierten en herramientas de carácter exponencial.

Hoy en día, los mayores logros obtenidos a través de las contribuciones de estas tecnologías están relacionados con las nuevas técnicas de aprendizaje que incluyen contenidos digitales que permiten asistir al profesor en muchas disciplinas (ODS4: Garantizar una educación de calidad inclusiva y equitativa y promover las oportunidades de aprendizaje a lo largo de toda la vida para todos); sin embargo, las aplicaciones que asisten al personal de salud en las intervenciones, mediante el guiado o la ayuda adecuada, han tenido un fuerte impacto en el ámbito sanitario (ODS3: Asegurar vidas saludables y promover el bienestar de todos en todas las edades). Finalmente, la Industria 4.0 [2] tiene en la RV/RA un sinfín de aplicaciones que permitirían trasladar las fábricas más avanzadas a lugares con alto índice de pobreza, con lo que se tiene una ayuda importante para la consecución del ODS8, relacionado con el Empleo y el Crecimiento Económico.

Actualmente, la pandemia de COVID-19 ha acentuado aún más las luchas geopolíticas, la crisis ambiental y la desigualdad y exclusión de los grupos más vulnerables. En el aspecto sanitario, el COVID-19 agravó el sufrimiento de las personas con dolencias mentales y físicas. El sistema sanitario nunca se había enfrentado a un enemigo de estas características, lo que ha obligado a buscar soluciones innovadoras que hagan que la sanidad mundial sea más flexible y esté preparada para el futuro ante este tipo de perturbaciones.

Las investigaciones sobre epidemias pasadas han revelado una profunda y amplia gama de consecuencias físicas y psicológicas durante los brotes, a nivel individual y comunitario, las cuales se han ratificado en la actual pandemia. Desde el punto de vista psicológico, existen múltiples alteraciones asociadas que van desde

síntomas aislados a trastornos complejos con un marcado deterioro de la funcionalidad. Desde el punto de vista físico, es notable el aumento de dolencias asociadas al aislamiento, el home office y el cambio en el estilo de vida. Con la declaración oficial de la pandemia por parte de la Organización Mundial de la Salud, las políticas públicas ahora apoyan el financiamiento de proyectos de investigación e innovación orientados a atender los desafíos sanitarios generados por la COVID-19. Asimismo, las empresas, universidades y centros de investigación relacionados con temas de salud y otros sectores de la economía reorientaron sus esfuerzos en materia de investigación y producción para hacer frente a la misma. Principalmente, la atención de los mercados ha girado hacia la RV/RA como herramientas potenciales que ayuden a afrontar las consecuencias derivadas de la pandemia.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Desde la gestión emocional en el día a día hasta los trastornos mentales, neurológicos o del neurodesarrollo, la “salud mental” es indispensable para nuestro bienestar, y sin embargo hasta antes de la pandemia era una de las grandes olvidadas. La Organización Mundial de la Salud define a la salud mental como un “estado completo de bienestar físico, mental y social” [3].

Desde las TIC, en el área de la salud física se ha trabajado en rehabilitaciones de pacientes que permitan el monitoreo y seguimiento del avance en su recuperación, y en el área de salud mental, se ha avanzado en la investigación de la detección temprana de trastornos, la mejora en la personalización de las intervenciones y el apoyo a un mayor porcentaje de la población. Como regla general, los sistemas sanitarios actuales deben: ser fáciles de usar, lograr una mayor implicación por parte del paciente y adaptarse mejor a sus necesidades.

Bajo la consideración del ODS3 es posible organizar las acciones de investigaciones

futuras en tres grandes líneas en las cuales las realidades alternativas pueden dar soporte:

- **Salud física:** Pretende poner recursos a disposición de las personas de forma que puedan recibir tratamiento a dolencias físicas. En este sentido, las realidades alternativas son un buen mecanismo de soporte que posibilitan la recreación virtual de terapias físicas como la rehabilitación de las capacidades motoras causadas por el ACV (granularidad fina), o la mejora de la resistencia, el equilibrio, la coordinación y el fortalecimiento muscular los cuales, por ejemplo, son síntomas comunes en la Esclerosis Múltiple (granularidad gruesa), entre otros [4,5,6,7,8].
- **Salud mental:** Primeramente, sin salud mental no hay salud, así que resulta evidente su relación con el ODS3. Una cobertura universal de salud debe incluir también la salud mental. El tratamiento de problemas de salud mental como la ansiedad, las fobias, los ataques de pánico, la depresión, la esquizofrenia, la psicosis, los trastornos alimentarios, los trastornos obsesivo-compulsivo y otros problemas relacionados con la mente, son grandes ejemplos de tratamiento médico donde las realidades alternativas pueden mejorar la vida de los pacientes. Por lo general, los pacientes son sumergidos en un entorno virtual seguro y controlado y enfrentados a sus trastornos [9,10,11,12].
- **Conciencia/Educación de Bienestar:** Esta línea permitiría completar el abordaje del cuidado de la salud propuesto por el ODS3; por medio del concientizar a la comunidad sobre la importancia de la buena salud y de un estilo de vida saludable. En este sentido, es posible ver que la pandemia ha posibilitado que la tecnología del mindfulness y el bienestar nos rodee, ya sea a través de

dispositivos de seguimiento de ejercicio hasta servicios de delivery saludables. Aunque hoy en día las tecnologías para el bienestar se utilizan con facilidad, por ejemplo, con accesorios de control del ritmo cardíaco, sesiones de terapia virtual y pruebas de biomarcadores, es evidente cómo el mundo del bienestar impulsado por la tecnología se ve reforzado por el crecimiento del mercado de la RA y la RV.

Entre las soluciones que proveen las realidades alternativas se encuentran aplicaciones para: monitorear el estado mental, yoga, asistencia a la meditación, promover la alimentación sana, educación sobre la información alimentaria de los envases, entre otras [13,14,15,16,17].

2. RESULTADOS OBTENIDOS/ESPERADOS

Las actividades realizadas hasta el momento por este grupo se han enmarcado dentro de 1 proyecto de Investigación de la UNSL, 1 proyecto de la Comunidad Europea, 4 proyectos de Desarrollo Tecnológico de la Secretaría de Políticas Universitarias y 3 proyectos de Extensión de Interés Social.

Como consecuencia del trabajo elaborado, se ha logrado desarrollar varios sistemas relacionados a RV con interacción multimodal involucrando aspectos verbales y gestuales, así como también se han elaborado aplicaciones de innovación y desarrollo asociadas a la RV y RA.

Actualmente las acciones se encuentran focalizadas en la incorporación de nuevas estrategias que permitan alcanzar una mejor percepción e interacción, al mismo tiempo que se logra incentivar al usuario tanto cognitiva como físicamente.

Además, se analizará el impacto logrado por el uso de sistemas de RV y RA en la adquisición y mejora de habilidades funcionales en procesos físicos y cognitivos asociados a las áreas de la salud y la educación, entre otras. De

esta manera, se pretende detectar y evaluar la evidencia científica resultante para determinar la envergadura de dichas intervenciones.

3. FORMACIÓN DE RECURSOS HUMANOS

Los trabajos realizados han permitido la definición de trabajos finales de carrera de la Licenciatura en Cs. de la Computación (4 finalizados), tesis de Especialización en Educación Superior (1 finalizada), tesis de Maestría en Cs. de la Computación (2 en ejecución y 1 finalizada) y tesis de Doctorado en Ciencias de la Computación (1 en ejecución).

Adicionalmente se ha obtenido una beca de iniciación a la investigación y una beca de perfeccionamiento en investigaciones otorgadas por la Secretaría de Ciencia y Técnica de la UNSL; y una beca doctoral de CONICET.

4. BIBLIOGRAFÍA

1. Lozano, H. E. B. La era del desarrollo sostenible: Nuestro futuro está en juego. Incorporemos el desarrollo sostenible a la agenda política mundial. *Revista Lumen Gentium*, 1(2), 101-103. 2017.
2. Basco, A. I., Beliz, G., Coatz, D., & Garnero, P. *Industria 4.0: fabricando el futuro* (Vol. 647). Inter-American Development Bank. 2018.
3. Lang, T., & Hauser, B. María Victoria Garrido Manzano Salud es un estado completo de bienestar físico, mental y social y no solamente la ausencia de afecciones o enfermedades (OMS). Libro de Resúmenes. 2017.
4. Lee, J. S. E., Chan, J. J. I., Ithnin, F., Goy, R. W. L., & Sng, B. L. (2020). Resilience of the restructured obstetric anaesthesia training programme during the COVID-19 outbreak in Singapore. *International journal of obstetric anesthesia*, 43, 89. 2020.
5. Goh, P. S., & Sandars, J. (2020). A vision of the use of technology in medical education after the COVID-19 pandemic. *MedEdPublish*, 9(49), 49. 2020.
6. Franchi, T. (2020). The impact of the Covid-19 pandemic on current anatomy education and future careers: A student's perspective. *Anatomical Sciences Education*, 13(3), 312. 2020.
7. Jofré, N., Rodríguez, G., Alvarado, Y., Fernández, J., & Guerrero, R. *Natural User Interfaces: A Physical Activity Trainer*. In *Argentine Congress of Computer Science* (pp. 122-131). Springer, Cham. 2017.
8. Jofré Pasinetti, N., Rodríguez, G., Alvarado, Y., Fernández, J., & Guerrero, R. A. El uso de la realidad virtual inmersiva en terapias motrices. In *XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, ITBA, Buenos Aires)*. 2017.
9. Brito C., Héctor, & Vicente P., Benjamín. (2018). Virtual reality and its applications in mental disorders: a review. *Revista chilena de neuro-psiquiatría*, 56(2), 127-135. 2018.
10. Deng, W., Hu, D., Xu, S., Liu, X., Zhao, J., Chen, Q., ... & Li, X. The efficacy of virtual reality exposure therapy for PTSD symptoms: A systematic review and meta-analysis. *Journal of affective disorders*, 257, 698-709. 2019.
11. L.V. Eshuis, M.J. Gelderen, Mirjam Zuiden, Mirjam Nijdam, Eric Vermetten, Miranda Olf, and A. Bakker. Efficacy of immersive ptsd treatments: A systematic review of virtual and augmented reality exposure therapy and a meta-analysis of virtual reality exposure therapy. *Journal of Psychiatric Research*, 11 2020.
12. Alvarado, Y., Rodríguez, G., Jofré, N., Fernández, J., & Guerrero, R. A. (2021). Post COVID-19 Cognitive disorders: Virtual Reality and Augmented Reality as mental healthcare tools. In *XXVII Congreso Argentino de Ciencias de la*

- Computación. 2021.
13. Ma, M., & Zheng, H. (2011). Virtual reality and serious games in healthcare. In *Advanced computational intelligence paradigms in healthcare 6. Virtual reality in psychotherapy, rehabilitation, and assessment* (pp. 169-192). Springer, Berlin, Heidelberg.
 14. Haixia, L., & Yongrong, W. (2020). Structural optimization of yoga top based on 3D virtual-reality technology. *The Journal of The Textile Institute*, 111(6), 916-923. 2020.
 15. Järvelä, S., Cowley, B., Salminen, M., Jacucci, G., Hamari, J., & Ravaja, N. (2021). Augmented virtual reality meditation: Shared dyadic biofeedback increases social presence via respiratory synchrony. *ACM Transactions on Social Computing*, 4(2), 1-19. 2021.
 16. Csakvary, B. Promoting healthier food choices with the application of Augmented Reality. 2017.
 17. Juan, M., Charco, J. L., García-García, I., & Mollá, R. An augmented reality app to learn to interpret the nutritional information on labels of real packaged foods. *Frontiers in Computer Science*, 1, 1. 2019.

QGIS Y BONOS VERDES: HERRAMIENTA PARA LA CLASIFICACIÓN DE PATRONES EN IMÁGENES GEOGRÁFICAS PARAMETRIZADAS SEGÚN EL CARBONO CAPTURADO POR LA BIOMASA FORESTAL.

Caffetti, Yanina A. (yaninacaffettimac@gmail.com) - Vargas, Alejandro R. (hidrogis2020@gmail.com)

Facultad de Ciencias Forestales - Universidad Nacional de Misiones.

Palabras claves: bonos verdes, economía verde, QGIS, imágenes satelitales.

RESUMEN

El presente artículo forma parte de un proyecto de investigación llevado adelante por docentes de la Facultad de Cs. Forestales, de la Carrera: Tecnicatura en Sistemas de Información Geográfica y Teledetección. El objetivo es ofrecer a la Provincia de Misiones una herramienta desarrollada en software libre en formato de capa para detectar especies nativas a través de imágenes satelitales manipuladas por QGIS, realizar un inventario para luego detectar patrones en imágenes aéreas capturadas por drones, con la finalidad de valorarlas y convertirlas posteriormente en bonos verdes.

CONTEXTO

En Escocia se realizó, durante el año 2021, la 26° Conferencia de las Naciones Unidas sobre el Cambio Climático (COP26). La Provincia de Misiones tuvo una representación propia en el evento, debido a la importancia de su patrimonio natural y sus políticas de protección al medio ambiente. En dicho encuentro se firmó un convenio con la empresa suiza Mercuria Energy Tardan que tiene vigencia hasta 2030 y enmarca la introducción de la Provincia de Misiones al mercado de bonos de carbono.

En este contexto, la línea de investigación presentada por medio de este artículo, pretende crear una herramienta con

software libre para detectar patrones en imágenes aéreas. Específicamente, valorar las imágenes de aquellas especies nativas que tengan mayor concentración de carbono capturado, dado que una tonelada de carbono (CO₂) equivale a un bono en el mercado verde.

1. INTRODUCCIÓN

En la actualidad, el cambio climático es la mayor amenaza medioambiental a la que se enfrenta el planeta. El aumento de la temperatura global, el deshielo de los polos, etc., son problemas que necesitan una inminente solución antes de que la situación sea irreversible.

La creciente preocupación por el cambio climático también está presente en el ámbito empresarial, donde actualmente, las “finanzas sostenibles” están de moda. Este término, según el Grupo de Expertos en Finanzas Sostenibles de la Unión Europea, supone la consideración de factores ambientales, sociales y corporativos en la toma de decisiones de las empresas.

De esta manera, al centrarse en los factores ambientales de las finanzas sostenibles, surge el término “finanzas verdes”, es decir, acciones o inversiones empresariales destinadas a reducir las emisiones de CO₂, promover las energías renovables, etc. Existen diversos instrumentos para la financiación de inversiones comprometidas con el medio

ambiente, siendo uno de ellos los conocidos como “bonos verdes”.¹

Santos Miranda, M. y Ferrera Rodríguez, R. (2017) definen este instrumento como “un tipo de bono que sirve para financiar proyectos de energías alternativas, eficiencia medioambiental, transporte limpio o adaptación al cambio climático”, pudiendo ser emitidos tanto por organismos públicos como por entidades privadas.² Por tanto, los bonos verdes funcionan exactamente igual que cualquier bono tradicional; el uso exclusivo de los fondos en inversiones medioambientales sostenibles es lo que diferencia a los bonos verdes del resto de bonos, siendo este un requisito imprescindible para que dicho bono sea calificado como verde, junto a la verificación y auditoría del mismo. Así, este instrumento financia desde parques eólicos o paneles solares hasta la gestión de bosques, residuos, eficiencia energética, recuperación de tierras, etc.

La primera emisión de bonos verdes se dio en el año 2007, cuando el Banco Europeo de Inversiones (BEI) utilizó este instrumento con el nombre de “Bono de Concienciación Climática”, por un importe de 600 millones de euros para financiar proyectos medioambientales. A esta emisión le siguieron otros bancos multilaterales de desarrollo como el Banco Africano de Desarrollo (BAFD) o el Banco Mundial, iniciando un mercado al que pronto se incorporaron emisores privados y entidades públicas, ampliando así el número de inversores y proyectos verdes.¹

La Provincia de Misiones, declarada Capital Nacional de la Biodiversidad, se encuentra en el noreste del país, en la frontera con Brasil y Paraguay, en lo que se conoce como mata atlántica, una formación vegetal que, en Argentina, sólo abarca a esa provincia. La mitad del suelo misionero, que sólo representa el 2% del territorio argentino, se encuentra bajo algún régimen de conservación y en ella habita el 52% de

la biodiversidad del país. Durante el 2021, y en la COP26, suscribió un convenio con Mercuria Trading que habilitó el proceso para que una empresa certificadora informe la cantidad de créditos de carbono que se genera en determinado territorio. La provincia entonces podrá acceder a los mercados de carbono y recibiría recursos monetarios a partir de los créditos de carbono generados entre 2017 y 2030 a precios de mercado, ingresos que podrían verse incrementados con la valorización en el tiempo de los créditos de carbono. La certificación se hará de acuerdo al marco jurisdiccional y de anidación REDD+ (JNR) de Verra³, organización que administra los estándares internacionales más utilizados en los mercados voluntarios de carbono. De esta manera inicia su participación en el denominado Mercado Internacional de bonos de carbono, generando bonos que produzcan beneficios económicos y de preservación de la biodiversidad.⁴

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

En el contexto actual, desde la Facultad de Ciencias Forestales, surge la hipótesis: ¿Es posible determinar a través de imágenes satelitales e imágenes aéreas la cantidad de CO₂ que absorbe la biomasa forestal? A través de modelos matemáticos se puede cuantificar el dióxido de carbono que absorbe una especie forestal, el desafío presentado es el diseño e implementación de una capa en QGIS con esas características. Para hallar el CO₂ en los árboles se debe entonces calcular la biomasa. Los inventarios forestales constituyen una base de partida para estimar la biomasa forestal.⁵ Para realizar el inventario, se planifica trabajar con imágenes satelitales manipuladas con el software QGIS⁶ que detecten monte nativo. Luego se trabajará con imágenes aéreas capturadas por drones en donde se identifique especies previamente clasificadas destacando aquellas que tengan mayor carbono capturado por la

biomasa forestal. Los pasos a seguir son los siguientes:

- Determinar la biomasa forestal por especies nativas, es decir el volumen de biomasa aérea que luego se determinará mediante un cálculo matemático el índice de biomasa en raíces.
- Luego del procesamiento inicial de datos contrastar los resultados con el índice creado por el Grupo Intergubernamental de Expertos sobre el Cambio Climático (IPCC) que facilita evaluaciones integrales del estado de los conocimientos científicos, técnicos y socioeconómicos sobre el cambio climático, sus causas, posibles repercusiones y estrategias de respuesta.⁷
- Crear un inventario en formato base de datos que contenga imágenes satelitales y aéreas de las especies, dichas imágenes serán procesadas en un primer momento por el software QGIS.
- Extraer los contornos de las imágenes aéreas y las características específicas por especie, que puedan servir de aprendizaje profundo a un modelo computacional compuesto por múltiples capas de procesamiento y representaciones de datos con múltiples niveles de abstracción.
- Generar una nueva capa plugin para QGIS con los datos procesados que se caracterice por la detección de patrones dentro de una imagen que correspondan a las especies con mayor captura de CO₂.
- Corroborar los resultados obtenidos realizando un relevamiento de especies nativas presentes en la Reserva de Uso Múltiple Guaraní (RUMG) de la Facultad de Ciencias Forestales.

- Calcular la cantidad de bonos verdes presentes en la RUMG para determinar un ejemplo de implementación.

3. RESULTADOS OBTENIDOS/ESPERADOS.

Naciones Unidas aprobó en 2015 la “Agenda 2030 sobre el Desarrollo Sostenible”. Este “plan de acción” global incluye diecisiete “Objetivos de Desarrollo Sostenible” (ODS) tales como la eliminación de la pobreza o la defensa del medio ambiente. El avance en el cumplimiento de estos objetivos requiere de un volumen significativo de inversiones por parte de gobiernos y empresas. Los Green Bonds o Bonos Verdes pueden ser una herramienta útil para canalizar recursos financieros hacia inversiones en infraestructuras que tengan un impacto positivo en el cumplimiento de estos objetivos; siendo en la práctica un recurso adecuado para llevar a cabo la transición hacia una economía más sostenible.⁸

En este escenario, a través de la carrera Tecnicatura Universitaria en Sistemas de Información Geográfica y Teledetección de la Facultad de Ciencias Forestales, se pretende diseñar y ofrecer una herramienta capaz de determinar la cantidad de bonos verdes presentes en la biosfera misionera, a través de imágenes satelitales y aéreas. El proyecto actualmente se encuentra en etapa de planificación.

4. FORMACIÓN DE RECURSOS HUMANOS.

El proyecto representa una de las líneas de investigación abordada por la doctorando en Informática Mgter. Yanina A. Caffetti así como también un eventual proyecto de investigación y extensión de la carrera Tecnicatura Universitaria en Sistemas de Información Geográfica y Teledetección de la Facultad de Ciencias Forestales.

5. BIBLIOGRAFÍA

1. Fernández Hinojosa, J. C. (2018). Los bonos verdes: el caso de Iberdrola.
2. Santos Miranda, M. y Ferrera Rodríguez, R. (2017). Bonos verdes: una forma de financiación en continuo crecimiento. *Estrategia Financiera*, 348, 42-48.
3. <https://verra.org/project/jurisdictional-and-nested-redd-framework/>
4. <https://economis.com.ar/como-sera-el-proceso-de-monetizar-los-bonos-verdes-de-misiones/>
5. BROWN S. (2002). Measuring carbon in forests: current status and future challenges. *Environmental Pollution* 116, 3, 363-372.
6. Graser, A. (2013). *Learning QGIS 2.0*. Packt Publishing Ltd.
7. <https://archive.ipcc.ch/index.htm>
8. Morjani, M. Y. E. (2018). Valoración de instrumentos emergentes de financiación: Bonos verdes.

IEI – Innovación en Educación en Informática

Aplicaciones Móviles, Realidad Aumentada y TVDi

María José Abásolo^{1,2} , Magdalena Rosado³ , Evelyn Del Pezo Izaguirre⁴ , María José Bouciguez⁵,
Wilma Gavilanes⁶, Florencia Puppo¹, Graciela Santos⁵, Cesar A. Collazos⁸ , Telmo Silva⁹ , Franco
Ronchetti¹, Cecilia Sanz¹ , Armando De Giusti¹ , Marcelo Naiouf¹ , Patricia Pesado¹ ,

¹ III-LIDI Instituto de Investigación en Informática,

Facultad de Informática – Universidad Nacional de La Plata (UNLP), Argentina

{mjabasolo, degiusti, mnaiouf, ppesado, csanz}@lidi.info.unlp.edu.ar

² Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CICPBA), Argentina

³ Facultad de Ingeniería Industrial, Universidad de Guayaquil, Ecuador

⁴ Universidad del Río, Ecuador

⁵ Facultad de Ciencias Exactas, Universidad Nacional del Centro de la Pcia. de Bs.As., Argentina

⁶ Universidad Técnica de Ambato, Ecuador

⁷ Facultad de Ciencias Humanas, Universidad Nacional del Centro de la Pcia. de Bs.As., Argentina

⁸ Departamento de Sistemas, Universidad del Cauca, Colombia

⁹ Universidad de Aveiro (UA), Aveiro, Portugal

Resumen

La línea de investigación y desarrollo presentada en este artículo realiza actividades de análisis, desarrollo y evaluación de aplicaciones móviles, videos interactivos, basadas en Realidad Aumentada y para TVDi. En particular se desarrollan aplicaciones para TVDi destinadas a adultos mayores. Por otra parte se estudian y desarrollan aplicaciones interactivas educativas destinadas a discapacitados auditivos. A la vez se estudia el uso y producción de videos interactivos con estudiantes universitarios. Por último se continúa el desarrollo y prueba de aplicaciones móviles que brindan información al ciudadano, en particular sobre la presencia de agentes peligrosos para la salud.

Entre los principales objetivos se destaca la formación de recursos humanos y el fortalecimiento de la investigación mediante el trabajo intergrupar entre diferentes instituciones nacionales y extranjeras.

Palabras Clave: Realidad Aumentada, Videos interactivos, Televisión Digital Interactiva, Discapacitados Auditivos, Adultos Mayores, Información al Ciudadano

Contexto

Las diferentes líneas de investigación están enmarcadas en diversos proyectos:

- Proyecto IDEAS-CICPBA “AMAPAS Aplicaciones Móviles para la Medición de

Agentes Peligrosos para el Ambiente y la Salud” (2019-2020) M.J.Abásolo.

- Proyecto "Computación de Alto Desempeño: Arquitecturas, Algoritmos, Métricas de rendimiento y Aplicaciones en HPC, Big Data, Robótica, Señales y Tiempo Real" (2018-2021) acreditado UNLP Programa de Incentivos, M. Naiouf.

- Proyecto "Metodologías, técnicas y herramientas de Ingeniería de Software en escenarios híbridos. Mejora de proceso" (2018-2021) acreditado UNLP Programa de Incentivos, P. Pesado

- Red de Aplicaciones y Usabilidad de la Televisión Digital Interactiva (RedAUTI) coordinada por la UNLP.

- Proyecto “PGTF INT/19/K08: TDTi*ABC La Televisión Digital Terrestre aplicada al mejoramiento de los pueblos de los países en vías de desarrollo” financiado por el Fondo Fiduciario Pérez Guerrero de la Oficina de Naciones Unidas para la Cooperación Sur-Sur, coordinado por la UNLP, se lleva a cabo con la CUJAE (Cuba) y la UFPB (Brasil).

Introducción

Desde hace varios años el grupo de investigación se dedica al desarrollo de aplicaciones en las áreas de Visión por Computador, Realidad Aumentada (RA) y Realidad Virtual (RV), con resultados publicados en ediciones anteriores de WICC [1-5].

Por otra parte, desde el año 2012 diversas universidades iberoamericanas conforman

RedAUTI, y el grupo de investigación ha presentado en ediciones anteriores de este workshop las líneas de investigación relacionadas con la Televisión Digital Interactiva (TVDi) [6-8]. En este artículo se presentan los principales avances alcanzados en esas diferentes líneas de investigación fruto de un trabajo conjunto con investigadores de diferentes instituciones del país en colaboración con el extranjero.

Aplicaciones móviles educativas para niños con discapacidad auditiva

Las Tecnologías de Información y Comunicación (TIC) adquieren protagonismo en el proceso de comunicación de las personas sordas al permitirles percibir las palabras a través de medios audiovisuales interactivos, disponibles en los dispositivos móviles y posibilitando su adaptación a diferentes contextos del entorno. La doctoranda Del Pezo [9-10] realiza una revisión sistemática de artículos sobre aplicaciones educativas, para niños sordos

Predomina la enseñanza de letras, vocabulario, pronunciación a partir del método reproductivo y como sistema de comunicación lidera la lengua de señas, lo que guarda relación directa con la enseñanza de la lectura, escritura y el habla al estar relacionados a partir de la fonología o el sonido de una palabra, que permite relacionarlas cognitivamente y visualmente para facilitar el proceso de aprendizaje buscando siempre fortalecer la lengua de señas como primer idioma.

La segunda metodología más utilizada es juegos serios con resolución de problemas y se orienta en mayor proporción al área de gramática buscando fortalecer en las personas sordas sus capacidades de comunicación para permitirles entender e interactuar con la sociedad oyente. Los juegos serios con resolución de problemas o reproductivos priorizan como técnicas lúdicas las estrategias de memoria, cognitivas combinadas con compensación y metacognitivas, ratificando que pueden utilizarse de manera complementaria y evolutiva a partir de la abstracción organizada del conocimiento, que posteriormente deberá relacionarse y aplicarse en contextos similares a los adquiridos o nuevos para evidenciar las destrezas en una determinada área, para que a través del sistema de recompensas automatizan al usuario a partir de sus logros obtenidos a continuar en el proceso de EA.

La mayoría de las aplicaciones móviles se desarrollaron para equipos con sistemas operativos Android que tienen mayor registro de

cuota de mercado, como canal audiovisual se mantiene el uso de elementos tradicionales multimediales como herramienta de apoyo de las aplicaciones, mientras que entre la RA y RV, lidera la RA al complementar su interacción también con multimediales.

Las herramientas de desarrollo tienden a ser de libre licenciamiento, sin embargo, hay incursiones en desarrollo híbrido (libre y de pago) sobre todo en aquellas aplicaciones en las que se desea mejorar la interfaz gráfica e integrar elementos con RE.

Existe un especial interés, y en este tema se enfoca la tesis doctoral de Del Pezo, en promover la enseñanza de la lectura de labios en las personas sordas de edad escolar apoyada en el uso de la tecnología..

Adultos mayores y TVDi

La inmersión de la tecnología en diversos ámbitos de la vida cotidiana ha permitido facilitar procesos que involucran la calidad de vida de la población mayor independiente. En [11-12] Rosado et al. presentan una revisión sistemática, desde 2012 hasta principios de 2019, de experiencias realizadas con personas mayores de 65 años, utilizando las TIC con el objetivo de mejorar su autocuidado y empoderamiento asistencial para el envejecimiento activo. La mayoría de los estudios analizados se refieren a aplicaciones móviles y servicios web que se centran en el seguimiento y asistencia a los adultos mayores, y una minoría centrados en la rehabilitación. Se encontraron algunos proyectos que utilizan la TVDi, y se enfocan en establecer recordatorios para atención médica y utilizar programas de entretenimiento para tener bienestar físico, mental y social.

La tesis doctoral de Rosado se enfoca en la promoción de la actividad física en adultos mayores mediante la TVDi. Debido a la familiaridad del uso de la televisión por parte del adulto mayor, en [13-14] Rosado et al. plantean el desarrollo de una aplicación con contenidos interactivos para TVDi que tenga por objetivo reducir las alteraciones de la marcha.

Se propone una evaluación inicial del adulto mayor para identificar si es independiente, y conocer la condición funcional de la marcha.

Para construir un programa de ejercicios, estos deben seguir ciertos parámetros como tipo/modo de actividad, duración, frecuencia, progresión, intensidad, densidad de entrenamiento. Los contenidos tendrán una progresión en la ejecución del ejercicio aumentando los niveles de dificultad, buscando cambios de velocidad y variedad de

ejercicios que involucren de cierta manera la diversión del adulto mayor. En este escenario los aspectos considerados para la ludificación del anciano, permitirá potenciar las habilidades que se requiere recompensar en la condición funcional de la marcha, lo que permitiría dinamizar su mejor desenvolvimiento y por consiguiente su calidad de vida.

Actualmente la aplicación se encuentra en desarrollo y se planea su evaluación con adultos mayores que asisten al centro gerontológico municipal “Dr. Arsenio de la Torre Marcillo” de la ciudad de Guayaquil.

Videos interactivos para la enseñanza universitaria

La educación universitaria en el contexto de pandemia se vio ante la necesidad de migrar las clases de un modelo presencial a otro que las simule lo más fielmente posible mediante videoconferencia. También tomó relevancia la enseñanza basada en videos y, asociado a esta estrategia, la reutilización de videos disponibles en la web o de los mismos docentes en otras instancias de enseñanza. Antes las dificultades de conectividad tanto estudiantes como profesores comienzan a grabar las videoconferencias. En estos registros se puede hallar valiosa información sobre el desarrollo de cursos como por ejemplo las interacciones entre docentes y estudiantes, puestas en común, explicaciones, exposiciones teóricas, aclaración de dudas, estrategias creadas “just in time” para resolver dificultades de aprendizaje puntuales, síntesis de aprendizajes creados de manera colaborativa por los estudiantes. Se consideró que sería posible transformar el material generado en clases videograbadas en videos con una capa interactiva que posibilite el diálogo cognitivo del estudiante con el contenido a aprender.

El objetivo del trabajo de Santos et. al. [15] es identificar prototipos de intervenciones pedagógicas para crear videos interactivos utilizando aplicaciones disponibles en la web. Se analizaron diferentes aplicaciones que permiten crear videos interactivos abordando funcionalidad y consideraciones didáctico-pedagógicas para diseñar actividades que se inserten como propuestas de interacción en los videos.

Incorporación de OCR a aplicación para brindar información de etiquetas nutricionales

El reconocimiento óptico de caracteres (OCR) es una técnica de identificación automática que se centra específicamente en la digitalización de textos. Se identifican los caracteres o palabras presentadas en una imagen en base a patrones conocidos. Las herramientas OCR permiten automatizar la introducción de información en un sistema informático. La posibilidad de realizar la carga automática mediante una herramienta de OCR en lugar de la carga manual por teclado realizada por una persona humana es una alternativa que permite una importante disminución del capital humano necesario para realizar la tarea, así como una suba en la productividad y una posible disminución en la tasa de errores junto al consecuente aumento en la calidad de la carga de datos.

En la tesina de grado de Puppo se tiene interés particular el reconocimiento automático de la información presente en etiquetas de productos nutricionales que circulan por la República Argentina. En este sentido el trabajo de Reibring [16] plantea un algoritmo de machine-learning para extraer información de etiquetas nutricionales. Destaca la ventaja de tener un cuadro estandarizado para todas las etiquetas de información nutricional de Estados Unidos, y la finitud del vocabulario utilizado para estos fines. A partir de estos dos aspectos, diseña un algoritmo de aprendizaje automático que es entrenado para reconocer palabras en lugar de caracteres.

Con el objetivo de proveer a los usuarios de una herramienta que les permita tomar decisiones más conscientes e informadas sobre los productos alimenticios que adquieren y consumen, Boza. y Prinsich [17] desarrollaron recientemente una aplicación denominada ConZoom que consiste en brindar la información nutricional de un alimento detallada en sus etiquetas de producto, a partir de la captura de su código de barras. Dicha aplicación se nutre de una base de datos que en la actualidad debe ser cargada manualmente. El objetivo de la tesina de Puppo es avanzar hacia una herramienta basada en OCR que pueda facilitar la carga semiautomática de la base de datos que almacena las características nutricionales de productos alimenticios.

Entre las implementaciones de OCR disponibles libres más conocidas se encuentran Tesseract OCR, GOCR, Cuneiform, OCRFeeder, Ocrad y ML Kit, las cuales se compararon reconociendo diferentes etiquetas nutricionales.

Líneas de investigación y desarrollo

- Aplicaciones móviles para brindar información y servicios al ciudadano sobre agentes peligrosos para la salud y el ambiente
- Aplicaciones móviles para enseñanza a niños con discapacidad auditiva
- Realidad Aumentada y Videos Interactivos aplicados a la educación universitaria
- VideoJuegos en la enseñanza de las ciencias
- Contenidos y Aplicaciones Interactivas para TVDi, con especial interés en asistencia a adultos mayores

Resultados y Objetivos

- Con el objetivo de formar recursos humanos desde el año 2012 se dicta la carrera de postgrado “Especialización en Computación Gráfica, Imágenes y Visión por Computadora”, Facultad de Informática de la UNLP. (N° 11.162/12).
- Se realizó un mapa de la ciudad de Tandil donde se ubicaron antenas y mediciones de radiaciones electromagnéticas de alta frecuencia. Por el momento el mapa puede ser visualizado en Google Maps¹. Se continúa trabajando en el desarrollo de una aplicación visualizadora de los datos georeferenciados basada en RA.
- Se está desarrollando una aplicación móvil lúdica para la enseñanza de lectura labial, a desarrollarse como parte de una tesis doctoral.
- Se organiza anualmente, desde 2012 a la actualidad, un evento científico que permite el intercambio de las investigaciones en el ámbito iberoamericano, denominado Jornadas de difusión y capacitación de Aplicaciones y Usabilidad de la TVDi. Se realizó la 10a Conferencia Iberoamericana de Aplicaciones y Usabilidad de la TVDi jAUTI 2021, Ecuador, 2-3 dic 2021.
- Se está desarrollando una aplicación para TVDi para asistir el mejoramiento de las alteraciones de la marcha en personas mayores, con el objetivo de probar los desarrollos realizados en Ecuador.
- Se realizó la comparativa de herramientas OCR como Tesseract OCR, GOCR, Cuneiform, OCRFeeder, Ocrad y ML Kit para el reconocimiento de caracteres de etiquetas nutricionales

- Se planifica la realización de cursos de doctorado conjunto con profesores de la UNLP y las instituciones extranjeras colaboradoras

Formación de recursos humanos

El equipo cuenta con un investigador formado trabajando en colaboración, para llevar a cabo la dirección conjunta de tesis junto con cuatro investigadores de universidades nacionales, España, Portugal y Colombia, de un tesista de grado y cinco tesistas de doctorado:

- Evelyn Del Pezo “Modelo colaborativo y lúdico basado en Aplicaciones Móviles para apoyar la enseñanza de la lectura labial a niños sordos” Director: Abásolo M.J. y Codirector Collazos C.A. (tesis de Doctorado en Ciencias en curso).
- Wilma Gavilanes “Metodología para la evaluación del impacto de experiencias con Realidad aumentada en educación superior” Director: Abásolo M.J. y Codirector J. Cabero (tesis de Doctorado en Ciencias en curso)
- María José Bouciguez “Ambientes virtuales altamente interactivos basados en videojuegos y simulaciones para la educación en ciencias” Directores: Santos, G. y Abásolo, M.J. (tesis de Doctorado en Ciencias, Fac. Inf. de la UNLP, finalizada pendiente de entrega)
- Magdalena Rosado “Televisión Digital Interactiva (TVDi) para reducir las alteraciones de la marcha en adultos mayores”. Directores: M.J. Abásolo y T. Silva (tesis de Doctorado en Ciencias UNLP en curso)
- Florencia Puppo “Reconocimiento óptico de caracteres de la etiqueta nutricional de productos alimenticios”. Directores: Abásolo M.J., Ronchetti F. (tesina de grado de Licenciatura en Sistemas, finalizada pendiente de entrega)

Además, se colabora en la formación de recursos humanos de otras universidades, entre los cuales se enumeran el becario doctoral de Conicet:

- Lucas Benjamin Cicerchia “Detección de enfermedades y falta de nutrientes en cultivos utilizando algoritmos de Active Learning aplicados al sensado remoto” Directores: Claudia Russo (UNNOBA), María José Abásolo (tesis de Doctorado en Ciencias UNLP en curso)

Referencias

¹ <https://www.google.com/maps/d/u/0/viewer?mid=12TAnlxBkzDtwqZ5pMGTrxmr72z-MSZEK>

- [1] Abásolo M.J., Gavilanes W., Del Pezo E., Allisiardi T., Bouciguez M.J., Santos G., Collazos C., De Giusti A., Naiouf M., Pesado P., Sanz C. *Realidad extendida y aplicaciones móviles aplicadas a la educación y a brindar información al ciudadano*. XXIII Workshop de Investigadores en Ciencias de la Computación WICC 2021, RedUNCI, ISBN 978-987-24611-3-3; 978-987-24611-4-0
- [2] Abásolo M.J., De Giusti A., Naiouf M., Pesado P., Barbieri S., Gavilanes W., Mitaritonna A., Vincenzi, M., Bria, O.; Ronchetti, F.; Montero F., Perales López F.; Springer, V. *Aplicaciones de realidad extendida y Aplicaciones Móviles*. XXII Workshop de Investigadores en Ciencias de la Computación WICC 2020, RedUNCI, ISBN 978-987-3714-82-5
- [3] Abásolo M.J., De Giusti A., Naiouf M., Pesado P., Sanz C., Barbieri S., Boza R., Gavilanes W., Mitaritonna A., Prinsich N., Vincenzi, M., Montero F., Perales López F. *Aplicaciones de realidad virtual, realidad aumentada e interfaces multimodales*. XXI Workshop de Investigadores en Ciencias de la Computación WICC 2019, Universidad Nacional de San Juan, RedUNCI, ISBN 978-987-3619-27-4
- [4] Abásolo M.J., Mitaritonna A., Castañeda, S., Sanz C., Boza R., Prinsich N., Silva, T., Rosado, M.; Naiouf, M.; Pesado, P.; De Giusti, A. (2018) *Aplicaciones de visión por computador, realidad aumentada y TVDi*. XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018, Universidad Nacional del Nordeste). p. 389-395, RedUNCI, ISBN: 978-987-3619-27-4
- [5] Abásolo, M.J.; Sanz, C.; Naiouf, M.; De Giusti, A.; Santos, G.; Castro, M.; Bouciguez, M.J. G. (2017) *Realidad Aumentada, Realidad Virtual e Interacción Tangible para la Educación*. XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017), pp. 1312-1316, RedUNCI, ISBN 978-987-42-5143-5
- [6] Abásolo M.J., Rosado Alvarez M., Silva T., Pina J., Socorro R., Kulesza R., Lemos G., De Giusti A., Naiouf M., Pesado P., *La Televisión Digital Interactiva para el mejoramiento de los pueblos latinoamericanos*. XXIII Workshop de Investigadores en Ciencias de la Computación WICC 2021, RedUNCI, ISBN 978-987-24611-3-3; 978-987-24611-4-0
- [7] Abásolo M.J., De Giusti A., Naiouf M., Pesado P., Rosado Alvarez M., Silva T., Pina J., Kulesza R. *Aplicaciones de la Televisión Interactiva y aplicaciones móviles para el mejoramiento de los pueblos latinoamericanos*. XXII Workshop de Investigadores en Ciencias de la Computación WICC 2020, RedUNCI, ISBN 978-987-3714-82-5
- [8] Abásolo M.J., De Giusti A., Naiouf M., Pesado P., Rosado Alvarez M., Pina J., Kulesza R., Silva T. *Aplicaciones de la Televisión Interactiva y tecnologías afines para el mejoramiento de los pueblos latinoamericanos*. XXI Workshop de Investigadores en Ciencias de la Computación WICC 2019, Universidad Nacional de San Juan, RedUNCI, ISBN 978-987-3619-27-4
- [9] Del Pezo Izaguirre, E., Abásolo, M. J., & Collazos, C. A. *Uso de tecnologías móviles y realidad extendida para personas sordas: Una revisión sistemática de la literatura de acceso abierto*. XV Conferencia Latinoamericana de Tecnologías de Aprendizaje.
- [10] E. Del Pezo Izaguirre, M. J. Abásolo and C. A. Collazos, *Educational Methodologies for Hearing Impaired Children Supported by Mobile Technology and Extended Reality: Systematic Analysis of Literature*. IEEE Revista Iberoamericana de Tecnologías del Aprendizaje, vol. 16, no. 4, pp. 410-418, Nov. 2021, doi: 10.1109/RITA.2021.3135202.
- [11] Rosado, M., Abásolo, M. J., & Silva, T. *ICT Oriented to the Elderly and Their Active Aging: A Systematic Review*. In Iberoamerican Conference on Applications and Usability of Interactive TV (pp. 134-155). Springer, Cham.
- [12] Rosado M., Abásolo M.J., Silva T. *Revisión de TIC Orientadas al Adulto Mayor y su Envejecimiento Activo*. En Abásolo M.J., Kulesza R. y Pina-Amargós J. (eds) Libro de aplicaciones y usabilidad de la televisión digital interactiva: jAUTI 2019. Universidad Nacional de La Plata.
- [13] Rosado, M., Abásolo, M. J., & Silva, T. *IDTV Application to Promote the Gait of the Elderly*. In Iberoamerican Conference on Applications and Usability of Interactive TV (pp. 134-146). Springer, Cham.
- [14] Rosado M., Abásolo M.J., Silva T. *Contenidos interactivos para TVDI destinados a reducir las alteraciones de la marcha en adultos mayores*. Proceedings of the 9th Iberoamerican Conference on Applications and Usability of Interactive TV - jAUTI 2020. Editorial UNLP, ISBN 978-972-789-680-6

[15] Santos G., Miranda A., Abásolo M.J. *Intervenciones pedagógicas para crear videos interactivos*. Proceedings of the 10th Iberoamerican Conference on Applications and Usability of Interactive TV - jAUTI 2021. Editorial UNLP (en prensa)

[16] Reibring, J.. (2017). *Photo OCR for Nutrition Labels. Combining Machine Learning and General Image Processing for Text Detection of American Nutrition Labels*. Gothenburg, Sweden: Master's Thesis Department of Signals and Systems Chalmers University of Technology, 2017

[17] Boza R. y Prinsich N. *Aplicación móvil para el consumo consciente y responsable*. Tesis de Grado Licenciatura en Sistemas, Facultad de Informática, 2019, UNLP. La Plata, Argentina.

Resultados preliminares en la formación de docentes y alumnos como Investigadores Científicos Iniciales

Bianchini Germán¹, Caymes-Scutari Paola^{1,2}, Ontiveros Patricia³, Rotella Carina³, Salinas Sergio⁴, Tagarelli Sandra⁴, Brancolini Alessandro¹, Chirino Pamela¹, Galdamez Mariela¹, Díaz Karvin¹

¹Laboratorio de Investigación en Cómputo Paralelo/Distribuido,
Departamento de Ingeniería en Sistemas de Información. Facultad Regional
Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza, +54 261 5244579

²Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

³Laboratorio de Gobierno Electrónico, Departamento de Ingeniería en Sistemas de
Información. Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza

⁴Laboratorio de Analítica de Datos, Departamento de Ingeniería en Sistemas de
Información. Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza

gbianchini@frm.utn.edu.ar, pcaymesscutari@frm.utn.edu.ar, pontiveros@frm.utn.edu.ar,
carinarotella@gmail.com, s4salinas@gmail.com, stagarelli@gmail.com,
alessandro.brancolini@alumnos.frm.utn.edu.ar, pamelaachirino@gmail.com,
mariela.galdamez.16@gmail.com, karvindiaz@gmail.com

RESUMEN

Junto con el inicio de la pandemia por COVID-19, que tantos cambios ocasionó en la actividad cotidiana y profesional de las personas, se inició el proyecto titulado 'Formación de docentes y alumnos de grado como Investigadores Científicos Iniciales en las áreas de Informática y Ciencias de la Computación', en el cual se buscó organizar y fortalecer el proceso de formación de investigadores en el Departamento de Ingeniería de Sistemas de Información de la UTN-Facultad Regional Mendoza, incluyendo a alumnos y docentes. En rigor, los responsables del mismo hemos venido realizando dicha tarea desde el año 2008, aunque sin el marco y formalidad de un proyecto marco

específico. En este artículo se comenta la razón del mismo y los resultados preliminares obtenidos tras la experiencia de este trabajo que concluirá a fines del actual ciclo 2022. Obviamente, las observaciones y conclusiones no pueden dejar de considerar la situación excepcional experimentada en el ámbito académico en relación al aislamiento y distanciamiento por COVID-19, y al impacto que las restricciones impuestas tuvieron en el desarrollo de los proyectos y las diversas actividades de investigación, también indirectamente afectadas por los cambios en la docencia.

Palabras clave: Investigación, Pandemia, Formación, Alumnos, Docentes, Proyectos.

CONTEXTO

El presente proyecto se realiza en el LICPaD (Laboratorio de Investigación en Cómputo Paralelo /Distribuido), y ha contado con la participación tanto de docentes del propio laboratorio, como de docentes de los grupos ADA-Lab (Laboratorio de Analítica de Datos) y GE-Lab (Laboratorio de Gobierno Electrónico), todos grupos pertenecientes al Departamento de Ingeniería en Sistemas de Información de la UTN-FRM. El proyecto cuenta con la acreditación, homologación y financiamiento de la Universidad Tecnológica Nacional a través del proyecto PID TEUTIME0007658TC. Mediante este trabajo se ha buscado conjugar la formación teórica y la práctica de investigadores iniciales para que éstos recorran todas las etapas necesarias que involucra el proceso científico en un ámbito de investigación. Para esto se planteó utilizar un trabajo de investigación específico como facilitador del proceso de aprendizaje, enfocado en las competencias específicas que deben adquirir los graduados de Ingeniería en Sistemas de Información, ya que al fomentarlas se contribuye también a incrementar el número y calidad de las investigaciones realizadas en la institución, además de impulsar el pensamiento crítico de docentes y alumnos frente a nuevos retos. En el caso del LICPaD, los participantes de la formación han sido alumnos/becarios, mientras que en el caso de ADA-Lab y GE-Lab, los participantes han sido los propios docentes.

1. INTRODUCCIÓN

La habilidad de investigar resulta un elemento imprescindible para hacer frente a los desafíos del mundo laboral. Sin

embargo, aunque la educación debiera estar orientada a generar pensamiento, la misma suele enfocarse en profesionalizar a los alumnos. Esta es una situación que debe modificarse, y puede ser un aporte de los propios docentes/investigadores, respaldados por la institución, abrir un nuevo panorama a los estudiantes y docentes universitarios en general, de modo tal de ayudarlos a desarrollar sus capacidades de evaluación, crítica e inventiva. La investigación es una actividad enfocada en obtener, mediante observación y experimentación, nueva información y conocimiento necesario para ampliar los diversos campos de la ciencia y la tecnología, sin perseguir, obligatoriamente alguna aplicación práctica inmediata.

En base a esto, es claro que la investigación podría considerarse como una de las actividades académicas más relevantes, lo que significa que debería contemplar la participación tanto de docentes como de estudiantes.

Existen múltiples casos de éxito de instituciones que apostaron a la formación científica de sus estudiantes, entendiendo a la investigación científica como eje fundamental para el desarrollo de una sociedad del conocimiento equitativa, viendo como resultado un mayor nivel y calidad en la educación.

El objetivo general de nuestro proyecto ha sido ofrecer una guía y brindar los elementos y acompañamiento que permitan estimular la formación integral en el campo de la investigación científica en un nivel y temática acorde a la capacidad, la preferencia, y formación previa de cada docente o alumno participante, teniendo en cuenta además para estos últimos el estadio en la carrera. Sin embargo, en general los estudiantes desconocen las actividades de investigación que se llevan a cabo en la institución precisamente porque este tipo

tareas transitan un camino que no se cruza con la vida académica cotidiana de los alumnos, situación que dificulta su acercamiento. A esto hay que sumar el aspecto económico relacionado con la realización de actividades de investigación. Indudablemente, es necesario que las instituciones cuenten con un importante apoyo y disponibilidad de presupuesto y recursos, tanto para remunerar el trabajo realizado y así incentivar la dedicación a investigación, como para brindar el entorno de trabajo equipado y adecuado a la temática que se aborde. Respecto de esto, es importante aclarar que encarar un proyecto de investigación científica normalmente es costoso, sobre todo considerando que los resultados suelen conseguirse a largo plazo, por lo que puede parecer una tarea poco atractiva. No obstante, es claro que, como inversión, resulta de gran importancia para el desarrollo de las instituciones y la sociedad.

La forma de trabajo para el presente proyecto, que debió adecuarse a las múltiples restricciones por la pandemia, fue: efectuar reuniones por videoconferencias mediante plataformas de videollamadas; realizar videos explicativos subidos a la plataforma Youtube que los alumnos pudiesen consultar todas las veces que fuese necesario; atender consultas a través de distintos canales (videoconferencias, foros, mails, WhatsApp, etc.).

En lo referido a la redacción, revisión y corrección de textos para la participación en eventos científicos, de transferencia y divulgación, se trabajó mediante iteraciones por mail, corrigiendo y reelaborando los borradores, mientras que la preparación de las exposiciones fue además supervisada mediante el uso de videoconferencia.

En lo atinente al acceso al clúster de computadoras de la Facultad, el mismo se

ha efectuado mediante acceso remoto, de modo que los alumnos becarios pudiesen hacer uso de dicho recurso sin la necesidad de trasladarse.

Finalmente, la asistencia a eventos se realizó de acuerdo a lo estipulado por los comités organizadores de cada uno, en general, coincidiendo la mayoría en la realización por modalidad virtual mediante distintas plataformas de videoconferencias.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Se abordaron diversas líneas temáticas, que en algunos casos fueron modificándose a lo largo del tiempo, organizadas en subgrupos de trabajo.

Dentro del área del Cómputo Paralelo, las líneas elegidas fueron: Paralelismo en algoritmos de aprendizaje para redes neuronales, Métodos de Inteligencia artificial para la reducción de incertidumbre en modelos de predicción de incendios, Redes neuronales paralelas aplicadas a la visión computacional, Predicción del avance del dengue en zonas endémicas y Multiplicación de Matrices.

Los docentes intervinientes del subgrupo ADA-Lab orientaron sus temas de investigación hacia Indicadores de procesos y medición, y los docentes del subgrupo GE-Lab primero se centraron en la temática de aplicación de Blockchain Federal (BFA) sobre la gestión de documentos judiciales, y luego en la aplicación de la tecnología de Firma Digital Remota en actas de exámenes de la UTN-FRM. En simultáneo se comenzó a esbozar un sistema para la autenticación facial de alumnos.

3. RESULTADOS OBTENIDOS/ESPERADOS

Como se adelantó en el Resumen del presente artículo, la pandemia trajo aparejada una serie de efectos negativos. Primero que nada, la falta de asistencia y respuesta por parte de los sectores de la Facultad/Universidad vinculados directa o indirectamente a investigación (Secretaría de Ciencia y Tecnología, Departamento de Compras y Suministros, Departamento de Ingeniería en Sistemas, etc.); la sobrecarga de tareas a raíz del esquema virtual; falta de recursos tecnológicos; dificultades para desarrollar trabajo grupal, dado que las reuniones virtuales no exhiben el mismo aprovechamiento de tiempo y dificulta ciertos aspectos en la comunicación de las personas, entre otras cosas, sin contar el problema adicional de no poder separar la vida familiar de la actividad laboral.

Sin embargo, también surgieron ventajas y oportunidades: el esquema virtual posibilitó el acceso a eventos que, por razones geográficas, de tiempo o costo, hubiesen sido imposibles de acceder. Además, hubo una gran oferta de cursos, webinars y demás, que también posibilitaron la participación desde el propio hogar. Otra ventaja ha sido la urgencia en incorporar herramientas que de otra forma no se hubiesen utilizado, o al menos no en tan poco tiempo.

En lo concerniente a los resultados concretos del proyecto, si bien la contribución principal que se ha buscado es la transferencia de formación por parte de docentes-investigadores hacia docentes y alumnos de grado y al medio educativo, también se ha impulsado que los alumnos aprendan y experimenten el trabajo en colaboración y por metas, dado que, en general, el trabajo de investigación no es algo que pueda realizarse de forma individual, puesto que normalmente requiere de espacios inter y transdisciplinarios.

Específicamente, dentro de los subproyectos, se han ido logrando avances que fueron presentados en el marco de eventos de ciencia y educación. Los primeros lineamientos de este trabajo y los temas vinculados fueron expuestos en el WICC 2020 mediante cinco artículos [1-5], un artículo en el XXVI CACIC [6], un artículo premiado como 3er mejor trabajo estudiantil en el 8° CONAISI [7] y otro artículo en E-ICES 15 [8].

Durante el ciclo 2021, se siguieron difundiendo aportes y avances con seis trabajos en el WICC 2021, cinco artículos en el IDETEC 2020 (evento que por la pandemia fue pospuesto hasta junio de 2021), un trabajo en el CADI, dos aportes en el 16 ICES, y tres artículos en el CONAISI 2021, cuyos trabajos aún se encuentran en prensa.

A su vez, los docentes investigadores, desde el comienzo del ciclo, han participado en diversos eventos y cursos para facilitar y mejorar la adaptación de la forma de trabajo y contar con más herramientas y experiencias que impulsen la concreción del objetivo global.

4. FORMACIÓN DE RECURSOS HUMANOS

En principio, el proyecto está pensado con alumnos-becarios, lo cual les permite dedicar cierta cantidad de horas semanales a la tarea de investigar de modo tal de llegar a cumplir los objetivos establecidos. No obstante, en tiempo de pandemia también los alumnos se han visto sobrecargados, por lo que en general escapan un poco a los compromisos que no sean un aporte concreto a la aprobación de las materias. Actualmente, el proyecto cuenta con dos becarios de grado y dos alumnos. A su vez, se cuenta con seis docentes. Tres son investigadores

formados con titulación doctoral, dos de los cuales cumplen el rol de guía del proyecto en general. El resto de los docentes se encuentra en formación, y aquí sucede un fenómeno similar al de los alumnos, pero por distintas razones. Los docentes que no poseen cargos específicos para investigación, no suelen dedicar mucho tiempo a las actividades científicas, lo que en definitiva demora los planes de trabajo y a su vez desalienta a los propios docentes ya que no ven resultados a corto plazo.

5. BIBLIOGRAFÍA

- [1] Bianchini G., Caymes-Scutari P., Ontiveros P., Rotella C., Salinas S., Chirino P., Galdamez M.: Formación de docentes y alumnos de grado como Investigadores Científicos Iniciales en las áreas de Informática y Ciencias de la Computación. In: Libro de Actas del XXII WICC 2020. ISBN 978-987-3714-82-5. pp.590-594 (2020)
- [2] Rotella C., Ontiveros P., Bianchini G., Caymes-Scutari P., Tagarelli S., Salinas S., Chirino P., Galdamez M.: Uso de la tecnología Blockchain Federal (BFA) para dejar pistas de auditoría y trazabilidad a sentencias y acordadas de la Suprema Corte de la provincia de Mendoza. In: Libro de Actas del XXII WICC 2020. ISBN 978-987-3714-82-5. pp.683-687 (2020)
- [3] Caymes-Scutari P., Bianchini G., Tardivo L., Méndez-Garabetti M.: Sintonización de Aplicaciones científico/ingenieriles: un proceso de medición-mejora para incrementar la eficiencia. In: Libro de Actas del XXII WICC 2020. ISBN 978-987-3714-82-5. pp.748-752 (2020)
- [4] Salinas S., Tagarelli S., Caymes-Scutari P., Bianchini G., Ontiveros P., Rotella C., Chirino P., Galdamez M.: Diseño de un sistema de medición de desempeño para Moodle en Educación Superior. In: Libro de Actas del XXII WICC 2020. ISBN 978-987-3714-82-5. pp.683-687 (2020)
- [5] Tagarelli S., Salinas S., Caymes-Scutari P., Bianchini G., Ontiveros P., Rotella C., Chirino P., Galdamez M.: Visualización de datos en un Tablero de Comando aplicado a plataformas de educación a distancia en el nivel Superior. In: Libro de Actas del XXII WICC 2020. ISBN 978-987-3714-82-5. pp.955-959 (2020)
- [6] Bianchini G., Caymes-Scutari P., Ontiveros P., Rotella C., Salinas S., Galdamez M., Chirino P., Diaz K., Ponce de León A.: Actividades de Investigación Científica con docentes y alumnos de grado en Informática y Ciencias de la Computación. In: Libro de Actas del XXVI CACIC 2020. ISBN 978-987-4417-90-9. pp.670-679 (2020)
- [7] Chirino P., Bianchini G., Caymes-Scutari P.: Análisis de Métodos de Inteligencia Artificial para la Reducción de Incertidumbre. In: Memorias de Trabajos 8vo CONAIISI. ISBN 978-950-42-0202-8. pp.1004-1011 (2020)
- [8] Caymes-Scutari P., Bianchini G., Tardivo M.L., Méndez M.: Modelo para acelerar la predicción de fenómenos de Métodos Evolutivo-Poblacionales. In: Actas de Trabajos Completos del E-ICES 15. ISBN 978-987-1323-66-1. pp.178-186 (2021)

Modelos, metodologías y recursos para el desarrollo del pensamiento computacional

Gladys N. Dapozo, Cristina L. Greiner, Raquel H. Petris, Yanina Medina, Ana M. Company, María C. Espíndola

Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura
Universidad Nacional del Nordeste. 9 de Julio N° 1449. Corrientes. Argentina.
{gndapozo, cgreiner, rpetris, yanina, acompa, mcespindola}@exa.unne.edu.ar

RESUMEN

Este proyecto busca contribuir con modelos, metodologías y recursos que contribuyan al desarrollo del pensamiento computacional, en la formación específica de profesionales de la Informática como también en la formación de estudiantes de otras disciplinas STEM y en la formación de formadores de los niveles educativos preuniversitarios. Para ello se proponen dos líneas de investigación: La promoción del pensamiento computacional en el nivel universitario y la formación de formadores de los niveles preuniversitarios. En la universidad, para mitigar indicadores de deserción y desgranamiento en estudiantes de Informática, y en formación docente de los niveles educativos preuniversitarios, para apoyar las políticas públicas orientadas a incorporar las Ciencias de la Computación en las escuelas, iniciativas que buscan incidir en un mayor desarrollo de las áreas STEM (*Science, Technology, Engineering and Mathematics*).

Palabras clave: Enseñanza de la programación. Formación de formadores. Pensamiento computacional.

CONTEXTO

Las líneas de I/D corresponden al proyecto PI 21F016 “Modelos, metodologías y recursos para el desarrollo del pensamiento computacional”, acreditado por la Secretaría de Ciencia y Técnica de la Universidad Nacional del Nordeste (SECYT-UNNE), continuación de las líneas de I/D del proyecto PI 16F016 (2018-2021).

1. INTRODUCCIÓN

Problemática que enmarca el proyecto

Existe un interés generalizado a nivel mundial para incrementar la cantidad y calidad de los

recursos humanos formados en disciplinas relacionadas con las ciencias, tecnologías, ingenierías y matemáticas (STEM). Diversos estudios [1][2] destacan la insuficiente producción de graduados en estas áreas, respecto de las necesidades del aparato productivo y, particularmente, la carencia de capital humano de TIC (Tecnologías de la Información y la Comunicación), con las capacidades necesarias para trabajar en la innovación digital.

En los últimos años, las instituciones educativas con formación en las ciencias de la computación y de la informática de los países más desarrollados (EEUU, Francia, España y otros países europeos), promueven el abordaje de los conceptos fundamentales de estas ciencias desde una visión renovadora, tal como lo manifiesta Wing: “*El Pensamiento Computacional será una habilidad fundamental usada por todos a mediados de Siglo XXI. A la lectura, escritura y aritmética, es necesario añadir el pensamiento computacional a la capacidad de análisis de cada estudiante*” [3].

Sumado a ello, cada ciencia se desarrolla, crece y avanza por el creciente influjo de la Informática, que contribuye al manejo de cantidades masivas de datos (i.e. Big Data) [4] permitiendo extraer conclusiones relevantes en cada área. Este fenómeno se observa aún en ciencias que hasta hace pocos años se habían mantenido al margen de estos desarrollos tecnológicos, como ser las ciencias sociales.

Pensamiento Computacional

El pensamiento computacional está relacionado con otros tipos de pensamiento: el matemático, el lógico y el crítico, entre otros, con los cuales comparte habilidades cognitivas comunes como ser reconocimiento de patrones, abstracción, modelado, repetición, entre otras.

Su finalidad es que a partir del reconocimiento de los aspectos que nos rodean, de problemas reales de las actividades diarias o de las ciencias, se propongan soluciones aplicando herramientas informáticas [5].

La tendencia mundial en educación es fomentar el pensamiento computacional y la enseñanza de la programación, desde etapas muy tempranas. Especialistas de diversos campos coinciden en que, habilidades como la creatividad, el análisis crítico, la colaboración y la comunicación cobrarán relevancia significativa, y que es necesario un cambio en el sistema educativo para lograr su desarrollo [6].

En nuestro país, acompañando esta corriente, se definieron políticas públicas que tienen como propósito cambios profundos en la enseñanza en escuelas primarias y secundarias en temas relacionados con las Ciencias de la computación [7]. Tanto la programación de computadoras como el desarrollo del pensamiento computacional resultan relevantes para este objetivo. Según Ripani [8] al comprender la lógica en la resolución de problemas, los alumnos también se preparan para entender y cambiar la realidad. La Fundación Sadosky lanzó en el año 2013 el proyecto Program.ar, que a partir de múltiples iniciativas acerca a niños y jóvenes al aprendizaje de las Ciencias de la Computación. En el 2015 el Consejo Federal de Educación declaró de importancia estratégica para el Sistema Educativo Nacional al aprendizaje de la programación durante la escolaridad obligatoria (Resolución CFE N° 263/15, 2015). En 2018, mediante la Resolución CFE N° 343/18 se aprobaron los Núcleos de Aprendizajes Prioritarios (NAP) de Educación Digital, Programación y Robótica. Entre sus objetivos se destacan los de fomentar el conocimiento y la apropiación crítica y creativa de las TIC, y desarrollar iniciativas orientadas a construir conocimiento sobre la programación y el pensamiento computacional.

Estas iniciativas conllevan la necesidad de formación de formadores capaces de encarar este proyecto en el ámbito de las escuelas en los distintos niveles.

Competencias, pensamiento complejo y pensamiento computacional

Por otra parte, en la Educación Superior, en el seno de las redes que nuclean las distintas profesiones (CONFEDI, RedUNCI, FODEQUI, entre otras) se debate sobre un nuevo paradigma de enseñanza y aprendizaje, centrado en el desarrollo de competencias.

Según Tobón [9] la formación de competencias se da desde el desarrollo y fortalecimiento de habilidades de pensamiento complejo como clave para formar personas éticas, emprendedoras y competentes. Este autor señala que se deben concebir las competencias como: *“Procesos complejos de desempeño con idoneidad en determinados contextos, integrando diferentes saberes (saber ser, saber hacer, saber conocer y saber convivir), para realizar actividades y/o resolver problemas con sentido de reto, motivación, flexibilidad, creatividad, comprensión y emprendimiento, dentro de una perspectiva de procesamiento metacognitivo, mejoramiento continuo y compromiso ético, con la meta de contribuir al desarrollo personal, la construcción y afianzamiento del tejido social, la búsqueda continua del desarrollo económico-empresarial sostenible, y el cuidado y protección del ambiente y de las especies vivas”*.

Edgar Morín [10], fundamenta que la complejidad y multi-disciplinariedad son parte del siglo XXI, por lo cual la educación del futuro debe estar centrada en la condición humana. Bajo este enfoque, el ser humano debe ser capaz de identificar los aspectos comunes y diversos que existen entre nosotros; y tener la capacidad de abstraer y descomponer un problema o contexto complejo en un conjunto de componentes simples que permitan mejorar los procesos de toma de decisiones y la resolución de problemas. Por ello, las universidades fomentan el desarrollo de las habilidades blandas (pensamiento crítico, trabajo en equipo, comunicación) igual de relevantes como las habilidades fuertes (técnicas) propias de la especialización.

El pensamiento computacional es un proceso mental que aplica habilidades propias de la computación y del pensamiento crítico. Cuenta

con cuatro principios: a) Descomposición de un problema en fases más pequeñas, b) Reconocimiento de patrones, c) Abstracción de información irrelevante al problema, d) Diseño de algoritmos para la resolución del problema. Similares principios proponen Selby [11], estableciendo también criterios para desarrollarlos: a) Generalización, b) Descomposición, c) Abstracción, d) Diseño Algorítmico, y e) Evaluación, como habilidad para reconocer y determinar los alcances de realizar procesos, en términos de eficiencia y uso de recursos.

En el proceso de enseñanza y aprendizaje de la programación se considera la necesidad de alcanzar en los estudiantes la comprensión del uso de la tecnología informática no como un simple requisito para aprender un lenguaje de programación mediante la elaboración de algoritmos y la generación de código, sino que se busca la apropiación de un proceso metodológico adecuado para resolver problemas y programar soluciones. El desarrollo de habilidades cognitivas como la capacidad de abstracción, una buena aptitud lógico-matemática y la facilidad para la resolución de problemas de orden algorítmico, son muy importantes al momento del aprendizaje de los fundamentos de programación [12].

Además, los vertiginosos cambios de los medios de comunicación y tecnológicos, han ido acompañados de profundas transformaciones en la vida personal y en las instituciones, y han puesto en crisis, entre otras cuestiones, la transmisión del conocimiento. Estos cambios marcan la necesidad de revisar algunas de las certezas que fuimos construyendo, especialmente desde los aprendizajes. Los fenómenos de big data, internet de las cosas, machine learning, inteligencia artificial, dan cuenta de la vertiginosidad del desarrollo tecnológico. En este contexto es valioso revisar qué propuestas vamos a generar en nuestras clases y que empecemos a dialogar con los sujetos reales que tenemos en las aulas para diseñar propuestas relevantes y profundas, definir qué es lo relevante en tiempos de información fluida

y accesible, qué los emociona, qué los fascina, qué los inquieta[13].

Es importante que los profesores asuman el compromiso de formarse en las particularidades que plantean los ambientes tecnológicos actuales, reconociendo que es fundamental abordar la tarea docente en forma colaborativa con otros colegas, pues la inteligencia humana opera en un modo distribuido. La utilización de cada innovación tecnológica impactará y moldeará el funcionamiento cognitivo de los estudiantes de manera decisiva, por lo cual la selección de los medios y recursos tecnológicos no es un tema que pueda soslayarse [14].

En este proyecto vinculado con procesos de enseñanza y TIC, no puede dejar de considerarse el impacto de la pandemia del Covid-19 en el sistema educativo. Si bien, la reacción inmediata, en todos los niveles, fue la de pasar, sin solución de continuidad ni preparación previa alguna, de la enseñanza presencial a una enseñanza remota para navegar en la emergencia, no se puede ignorar que la ausencia de presencialidad pudo haber afectado aprendizajes y profundizado las desigualdades. En la enseñanza pos pandemia tendrá seguramente una ponderable incidencia la estrategia de hibridación entre modalidades que se adopte, así como su alcance, intensidad, conveniencia pedagógica, y momento o fase del proceso de formación en que se introduzca o trabaje [15].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Para el logro de los objetivos de este proyecto se propone trabajar en las siguientes líneas:

1. Formular un modelo de formación de pensamiento computacional en la enseñanza de la programación en carreras de Informática.
 - 1.1. Diseñar y aplicar estrategias para desarrollar el pensamiento computacional de los estudiantes en la enseñanza de la programación en la formación de informáticos.
 - 1.2. Recabar información del proceso formativo como retroalimentación para evaluar el grado de adquisición de las habilidades cognitivas propias del

- pensamiento computacional por parte de los estudiantes.
2. Elaborar un modelo de formación de pensamiento computacional transferible a los docentes de los niveles educativos preuniversitarios.
 - 2.1. Desarrollar propuestas pedagógicas que involucren conceptos de programación y robótica educativa para trabajar en la escuela.
 - 2.2. Desarrollar actividades didácticas digitales que puedan ser utilizadas en distintas propuestas pedagógicas buscando un abordaje holístico para la resolución de problemas, aplicando conceptos de gamificación y juegos serios.
 - 2.3. Definir mecanismos de seguimiento y de comprobación del grado de adquisición de las competencias vinculadas con el PC por parte de los docentes
 - 2.4. Evaluar el grado de implementación de estrategias educativas vinculadas con el pensamiento computacional y las Ciencias de la Computación que los docentes que participan de las capacitaciones han realizado en sus espacios áulicos, mediante una encuesta semestral.
 3. Desarrollar un marco de referencia para evaluar las características de las herramientas visuales y/o lúdicas utilizadas en la enseñanza inicial de la programación para orientar a los docentes en la selección de las herramientas más adecuadas para incorporar en sus propuestas docentes
 - 3.1. Relevar y evaluar las herramientas visuales y/o lúdicas utilizadas en la enseñanza inicial de la programación.
 - 3.2. Definir criterios de evaluación de las herramientas, en función de aspectos tecnológicos como didácticos.
 - 3.3. Diseñar un marco de referencia que permita la selección de las herramientas en función de los criterios predefinidos y objetivos de aprendizaje
 - 3.4. Validar el marco de referencia con diferentes casos de estudios, diseñados para tal fin.
 4. Generar un repositorio digital de recursos (herramientas y actividades didácticas) para

los docentes de los niveles educativos preuniversitarios.

- 4.1. Realizar un relevamiento sobre recursos digitales orientados al desarrollo del pensamiento computacional en las escuelas.
- 4.2. Seleccionar y publicar recursos propios del grupo de investigación y otros recursos externos que resulten de interés, como así también las producciones de los docentes que asisten a los cursos de capacitación.
- 4.3. Generar espacios de intercambio y comunicación entre los diferentes participantes.
- 4.4. Poner a disposición el repositorio para los docentes de la región.
- 4.5. Evaluar el impacto y las visitas al sitio web con el propósito de optimizar los recursos disponibles.

3. RESULTADOS OBTENIDOS/ESPERADOS

En la línea de estrategias educativas para la enseñanza de programación en carreras de Informática, en [16] se presenta la experiencia del dictado en modalidad “remota de emergencia” realizada en Algoritmos y Estructuras de Datos I de la Licenciatura en Sistemas de Información de la FaCENA-UNNE. Se comentan las adecuaciones pedagógicas realizadas con el propósito de lograr los objetivos de formación, destacando los aspectos que resultaron favorables y los que requieren ajustes o nuevas estrategias.

En la línea de formación docente, en [17] se muestran los resultados del dictado de la Diplomatura en Programación y Robótica Educativa. Se destaca el interés que suscita esta formación por parte de los docentes, provenientes de distintas áreas de conocimiento y con una motivación en consonancia con los objetivos de esta propuesta educativa. Así también se comprueba que los docentes han incorporado los conocimientos y las técnicas propias de la programación y han consolidado competencias para la cultura digital, tales como la colaboración y la creatividad, necesarias para lograr en los alumnos un aprendizaje más efectivo y motivador.

4. FORMACIÓN DE RECURSOS HUMANOS

En este proyecto participan seis docentes investigadoras, entre ellas dos tesis de posgrado que desarrollan su Trabajo Final de la Maestría en Tecnologías de la Información de la UNNE.

5. REFERENCIAS

- [1] M. Borchardt, I. Roggi, *Ciencias de la computación en los sistemas educativos de América Latina*, Buenos Aires: IPE – UNESCO Cuaderno SITEAL. 2017.
- [2] R. Katz, “TIC, digitalización y políticas públicas,”. *Entornos Digitales y Políticas Educativas*. IPE-UNESCO, pp. 17-58. 2016.
- [3] J. Wing, “Computational Thinking,” *Communications of the ACM*, vol 49(3), pp. 33-35, 2006.
- [4] G. Simari, “Los fundamentos computacionales como parte de las ciencias básicas en las terminales de la disciplina Informática”, VIII Congreso de Tecnología en Educación y Educación en Tecnología. Santiago del Estero, 2013.
- [5] H. Pérez Narváez y R. Roig-Vila, “Entornos de programación no mediados simbólicamente para el desarrollo del pensamiento computacional. Una experiencia en la formación de profesores de Informática de la Universidad Central del Ecuador”, *RED-Revista de Educación a Distancia*, 46(9), 2015.
- [6] Arias Ortiz, E., y Cristiá, J. (2014). El BID y la tecnología para mejorar el aprendizaje: ¿Cómo promover programas efectivos? Nota Técnica IDB-TN-670, BID.
- [7] Sadosky. CC-2016 Una propuesta para refundar la enseñanza de la computación en las escuelas argentinas. 2013. Disponible en: <http://www.fundacionsadosky.org.ar/wp-content/uploads/2014/06/cc-2016.pdf>
- [8] M. F. Ripani. *Competencias de Educación Digital*. Ministerio de Educación de la Nación. Argentina. 2017. Libro digital. Disponible en: <http://www.bnm.me.gov.ar/giga1/documentos/EL005452.pdf>
- [9] S. Tobón, “El enfoque complejo de las competencias y el diseño curricular por ciclos propedéuticos”. *ACCIÓN PEDAGÓGICA*, N° 16, Enero - Diciembre, 2007 - pp. 14 - 28
- [10] E. Morin, *Los siete saberes necesarios para la educación del futuro*. Buenos Aires: Nueva Visión. 2002
- [11] C. C. Selby, *Relationships: computational thinking, pedagogy of programming, and Bloom's Taxonomy*. *WiPSCE '15 Proceedings of the Workshop in Primary and Secondary Computing Education*. 2015. (págs. 80-87). London, United Kingdom: ACM New York, NY, USA. doi:10.1145/2818314.2818315
- [12] J. A. Jiménez-Toledo, C. Collazos, O. Revelo-Sánchez, “Consideraciones en los procesos de enseñanza-aprendizaje para un primer curso de programación de computadores: una revisión sistemática de la literatura”. *Instituto Tecnológico Metropolitano. TecnoLógicas*, vol. 22, 2019.
- [13] C. Lion. *Aprendizaje y tecnologías. Habilidades del presente, proyecciones del futuro*. Centro de Publicaciones Educativas y Material Didáctico. 2020.
- [14] S. Coicaud. *Potencialidades Didácticas de la Inteligencia Artificial. Mediaciones Tecnológicas para una enseñanza disruptiva*. Centro de Publicaciones Educativas y Material Didáctico. 2019.
- [15] Eduardo Sánchez Martínez. ¿Qué universidad después de la pandemia?. *Integración y Conocimiento: Revista del Núcleo de Estudios e Investigaciones en Educación Superior de Mercosur*, ISSN-e 2347-0658, Vol. 11, N° 1, 2022, págs. 147-152
- [16] G. Dapozo, C. Greiner, A. M. Company, S. Armana. *Desarrollo de habilidades cognitivas para la programación en contexto de masividad y enseñanza remota*. Primeras Jornadas de Educación y TIC de FaCENA – UNNE. Universidad Nacional del Nordeste. Facultad de Ciencias Exactas. 2021. Libro Digital. ISBN 978-987-3619-71-7.
- [17] G. N. Dapozo, C. L. Greiner, R. H. Petris, Y. Medina, A. M. Company and M. C. Espíndola, "Motivación y logros en la formación de docentes para introducir la programación y la robótica en los niveles educativos no universitarios," 2020 IEEE Congreso Bienal de Argentina (ARGENCON), 2020, pp. 1-8, doi: 10.1109/ARGENCON49523.2020.9505572.

Didáctica del Pensamiento Computacional y la Programación en distintos Niveles Educativos

Ana Casali^{1,2}, Claudia Deco^{1,3}, Natalia Colussi¹, Pamela Viale^{1,3},
Cristina Bender^{1,3}, Natalia Monjelat⁴

¹Facultad de Ciencias Exactas, Ingeniería y Agrimensura,
Universidad Nacional de Rosario, Rosario, Argentina.
{acasali, deco, colussi, pamela, bender}@fceia.unr.edu.ar

²Centro Internacional Franco Argentino de Ciencias de la Información y de Sistemas
(CIFASIS: CONICET-UNR)

³Facultad de Química e Ingeniería del Rosario,
Universidad Católica Argentina, Campus Rosario, Argentina.

⁴Instituto Rosario de Investigaciones en Ciencias de la Educación
(IRICE: CONICET-UNR)

Resumen

Se presenta una línea de investigación sobre la didáctica de la Programación en distintos niveles educativos, con especial atención a la formación de formadores. Las estrategias utilizadas combinan un enfoque desconectado (unplugged) con actividades que implican la programación de computadoras y otros dispositivos. Por un lado, se presenta el trabajo de formación de docentes del nivel primario en la Programación y su Didáctica, que se realiza en la Regional VI de Santa Fe (Argentina). En relación con el nivel universitario, se presenta por un lado un proyecto sobre la enseñanza de la primer materia de programación a estudiantes de la Licenciatura en Ciencias de la Computación (LCC), Licenciatura en Matemática (LM) y Profesorado de Matemática (PM) en la Facultad de Ciencias Exactas, Ingeniería y Agrimensura (FCEIA-UNR). Además, se presentan experiencias desarrolladas en los cursos iniciales de carreras de ingeniería en la Facultad de Química e Ingeniería del Rosario (UCA). En todas estas experiencias se focaliza en la didáctica de la enseñanza y aprendizaje de la Programación y el Pensamiento Computacional.

Palabras clave: Programación, Pensamiento Computacional, Carreras STEM, Formación Docente nivel Primario y Universitario.

Contexto

Esta línea de I+D se está llevando a cabo a través de varios proyectos que comparten algunos de sus miembros docentes e investigadores, conformando una red de colaboración:

- PID-UNR 80020180300062UR “Aprendizaje y Enseñanza de las Ciencias de la Computación en el Nivel Primario” Directora A. Casali (2019-2022).
- PID-UCA “Pensamiento Computacional Aplicado a Educación” Directora C. Deco (2019-2022).
- Proyecto de Cooperación Fundación Sadosky y UNR para el dictado de cursos “La programación y su Didáctica” orientados al Nivel Primario. Coordinadora A. Casali (2021-2022).
- PID-UNR 80020190100255UR “Estrategias Didácticas para el Aprendizaje y la Enseñanza del Pensamiento Computacional en el

Nivel Académico Universitario”.
Directora N. Colussi (2020-2021).

Introducción

En los últimos años hay consenso internacional acerca de la gran relevancia de introducir la enseñanza y aprendizaje de las Ciencias de la Computación (CC) en todos los niveles de escolaridad. Es importante que los estudiantes en formación comprendan desde temprana edad, cómo funcionan las computadoras, cómo procesan la información y que puedan trabajar en la formulación de problemas y en expresar la descripción de su solución de forma que un computador pueda ejecutarla. Este modo de pensar se denomina Pensamiento Computacional (PC) y Wing [1] plantea que su aprendizaje beneficia no sólo a los futuros informáticos, sino a toda la sociedad y por lo tanto, su enseñanza debería estar presente en todos los niveles educativos.

Por otra parte, se presenta el problema de tener pocos estudiantes en las carreras vinculadas a las áreas de ciencia, tecnología, ingeniería y matemáticas, usualmente denominadas STEM, y el problema de la alta deserción en los primeros años de estas carreras. Por esto, las universidades deben buscar diferentes estrategias para que el alumnado pueda motivarse y tener mejores herramientas para continuar el programa de formación al cual pertenece; por ejemplo, estrategias para desarrollar las destrezas del PC en los estudiantes.

En el contexto europeo es posible relevar diferentes estudios donde se analizan procesos y situaciones vinculadas a la enseñanza de las CC, la programación y el PC, mientras que en el contexto latinoamericano se observan iniciativas aisladas, siendo necesario implementar políticas públicas al respecto [2, 3].

En nuestro país, el Consejo Federal de Educación señaló al aprendizaje de la programación de importancia estratégica para el Sistema Educativo Nacional durante la escolaridad obligatoria (Res. CFE N° 263/15, 2015) y en septiembre de 2018 se aprobaron los “Núcleos de Aprendizajes Prioritarios de Educación Digital, Programación y Robótica” (NAP EDPR) para los diferentes niveles de la educación obligatoria (Res. CFE N° 343/18, 2018). En este sentido, es posible señalar que desde el marco institucional existen dos aspectos clave para introducir la enseñanza de las CC en el sistema educativo argentino. Por un lado, la necesidad de realizar una modificación curricular y por otro, la implementación de la formación docente en esta área. En particular, en el contexto de la provincia de Santa Fe, la formación docente se ha enfocado en el uso de TICs en la práctica educativa.. Teniendo en cuenta esta realidad, se consideró importante desde la universidad llevar adelante un proyecto de formación de docentes del nivel primario en CC atendiendo a las particularidades del contexto.

A nivel universitario se trabajó en dos proyectos. Por un lado el cambio del paradigma de aprendizaje, enseñanza y evaluación a través de la aplicación del Aprendizaje Basado en Proyectos y Problemas (ABP) en el aula del Redictado de Programación I para las carreras universitarias LCC, LM y PM. Por otra parte, se introdujeron conocimientos de PC en la materia Informática dictada en los primeros años de las carreras de Ingeniería, ya que el proceso de pensamiento involucrado puede ayudarlos a resolver mejor los problemas inherentes a sus profesiones.

Líneas de Investigación y Desarrollo

Formación Docente de Nivel Primario en Didáctica de las CC

En esta dirección, a partir de un primer Convenio suscripto entre la Fundación Sadosky, la Universidad Nacional de Rosario y el Ministerio de Educación de la provincia de Santa Fe se diseñó la “Especialización Docente de Nivel Superior en Didáctica de las Ciencias de la Computación: Aprendizaje y Enseñanza del Pensamiento Computacional y la Programación en el Nivel Primario” (Res. ME N° 1565/17). La primera cohorte se dictó en el Instituto Superior de Formación Docente N° 36 "Mariano Moreno" de la ciudad de Rosario entre 2017 y 2019, con una carga horaria total de 400 hs, con el 80% presencial. Esta formación tuvo por objetivo formar docentes capaces de experimentar y reflexionar críticamente acerca del PC y la programación, a los fines de construir las competencias adecuadas al nivel primario que posibiliten una práctica educativa innovadora [4]. Este postítulo combinó los enfoques unplugged y plugged en sus diferentes módulos [2, 5] alrededor de los ejes: el PC, la programación y conceptos tecnológicos. Se incluyeron también proyectos integradores, con el fin de que faciliten a los maestros llevar al aula estos aprendizajes [6, 7]. La propuesta pedagógica se basó principalmente en el Aprendizaje Basado en Proyectos y Problemas, siguiendo el enfoque de Tecnología para la Inclusión Social

A través de un segundo Convenio de Cooperación entre la Fundación Sadosky y UNR para el dictado de cursos “La programación y su Didáctica” orientados al Nivel Primario, en el segundo cuatrimestre de 2021 se ha dictado el Curso del Nivel I

y en el próximo cuatrimestre de 2022 se dictará el curso de Nivel II. Se destaca el gran interés que ha habido en la inscripción de este primer curso, quedando muchos maestros en lista de espera y en tratativas con el Ministerio para que esta formación se pueda extender a toda la provincia.

Estrategias Didácticas para la Enseñanza del PC y la Programación en el Ciclo Inicial Universitario en Ciencias Exactas

Esta línea de investigación busca desarrollar e indagar sobre la aplicación de diferentes estrategias didácticas que permitan abordar las problemáticas de la enseñanza, aprendizaje y evaluación de los estudiantes recursantes de las materias introductorias a la programación en las carreras de LCC, LM y PM de la FCEIA-UNR. Para lograr esta transformación se rediseñó de una forma innovadora la presentación, ejercitación y evaluación de contenidos propios de la materia. Se utilizó la estrategia de Aprendizaje Basado en Proyectos y Problemas (ABP) realizando a lo largo del cursado dos proyectos grupales que permitieran a los estudiantes cambiar su rol dentro del cursado, lograr un aprendizaje significativo, creativo, curioso y autónomo [8]. Durante la pandemia del Covid19 se utilizó además, para la evaluación una vidriera de exposición de proyectos¹ donde se recopiló los videos expositivos de cada uno de los grupos de estudiantes [9]. Para medir la receptividad de los estudiantes de la forma de aprendizaje elegida utilizamos cuestionarios grupales entregados junto al informe de cada proyecto. También se registraron opiniones y manifestaciones espontáneas que los estudiantes nos hicieron llegar mediante correo electrónico y grabaciones hechas en las defensas del

¹ <https://sites.google.com/view/vidriera-proyectos-fceia/>

segundo proyecto [8, 9].

Experiencias del desarrollo del PC en Carreras de Ingeniería

Este proyecto, en el ámbito universitario, pretende incorporar el PC en el proceso de enseñanza y aprendizaje en las materias de Informática que se dictan en el primer año de las carreras de Ingeniería (UCA Campus Rosario), considerando que este proceso de pensamiento involucrado puede ayudar a los estudiantes a resolver mejor los problemas relacionados a sus carreras. Para esto se desarrolló material específico haciendo hincapié en aquellos conocimientos relacionados al PC, como ser: abstracción, descomposición de problemas en subproblemas, reconocimiento de patrones y algoritmos.

Para ver la efectividad de este entrenamiento, se aplicó a los alumnos, al inicio y al final del cursado, un test propuesto por Marcos Román-González [10, 11].

Resultados

Se ha diseñado e implementado una formación superior para docentes del nivel primario en CC, la primera en la provincia de Santa Fe. Los resultados son sumamente alentadores, tanto desde la capacitación de los maestros (31 maestros egresados) como del diseño de la propuesta curricular de la Especialización, aprobada por el Ministerio de Educación de la provincia de Santa Fe. Las opiniones docentes en cuanto a la didáctica y contenidos de distintos módulos han sido altamente positivas y pueden verse en [6]. Destacamos además, que en este proceso se ha conformado un equipo docente universitario e interdisciplinario, competente en la formación docente en CC para el nivel primario.

A partir de este equipo, y continuando con esta línea de acción, en el cuatrimestre pasado se ha dictado el Curso “La

Programación y su Didáctica I” a 101 maestros de Rosario y la Región VI, con cerca del 60% de aprobación y muy buenos resultados del impacto de este curso en sus prácticas docente.

En cuanto al ciclo inicial universitario en ciencias exactas, los estudiantes manifiestan en los cuestionarios realizados los aspectos positivos que les dejó el desarrollo de los proyectos grupales de programación. Además, destacaron cómo éstos los ayudaron a incorporar conocimiento disciplinar, mantenerse motivados durante todo el proceso de desarrollo del trabajo, y cómo la actividad grupal facilitó relacionarse con el resto de los compañeros fortaleciendo de esta forma el vínculo entre pares. Esta combinación permitió que los cursantes logaran regularizar y aprobar la materia con una apreciación sólida y un alto grado de confianza sobre los saberes disciplinares adquiridos.

Con respecto a la incorporación del PC en el primer año de carreras de ingeniería, se analizaron las diferencias entre las respuestas al test de Marcos Román-González, al inicio y final de los cursos de Informática. Esta comparación nos brindó valiosa información acerca de la aprehensión de los conocimientos sobre el PC por parte de los alumnos y dio una pauta sobre la calidad del material desarrollado. Se trabaja sobre los contenidos de los apuntes de estas cátedras, considerando estos resultados.

Formación de Recursos Humanos

El equipo de trabajo está integrado por las doctoras Ana Casali, Claudia Deco, Pamela Viale y Natalia Monjelat, la magister Cristina Bender, la licenciada Natalia Colussi y el estudiante Hernán Galardi. Dentro del marco de esta línea de I+D se desarrolló el trabajo de doctorado en Ciencias de la Educación de la Ing. Maritza

García Angarita, cuya defensa se realizó en diciembre de 2022 y algunos resultados pueden verse en [12].

Cabe destacar además, la formación que se lleva a cabo de docentes del Departamento de Ciencias de la Computación en la enseñanza de la disciplina a docentes de nivel primario y secundario, es un trabajo interdisciplinario con especialistas de las ciencias de la educación.

Referencias

- [1] Wing, J. Computational Thinking Benefits Society. *Social Issues in Computing*. 2014.
- [2] Brackmann C., Román-González M., Moreno-León J., Robles G., Casali A. and Barone D. Computational Thinking Unplugged: Teaching and Student Evaluation in Primary Schools. In WIPSCENijmegen, The Netherlands, ACM. November, 2017.
- [3] Brackmann C., Couto Barone D., Casali A., Román-González M., Panorama Global Da Adoção Do Pensamento Computacional (Cap.3) Computação na educação básica: fundamentos e experiências (Ed. Raabe, Zorzo e Blikstein), pp. 30-47, 2020.
- [4] Monjelat, N. "Programming Technologies for Social Inclusion," En LACLO, 2017. La Plata, Argentina. 2017.
- [5] Bell, Witten & Fellows. CS Unplugged: An enrichment and extension programme for primary-aged students. 2015.
- [6] Casali A., Zanarini D., Monjelat N & San Martín P., Primary Level Teachers Training in Computer Science: Experience in the Argentine Context, In: Pesado P., Arroyo M.(eds) Computer Science—CACIC 2019. CCIS 1184, pp. 389-404, Springer 2019 https://doi.org/10.1007/978-3-030-48325-8_25
- [7] A. Casali, P. San Martín, N. Monjelat, P. Viale, Experiencias y aprendizajes del trayecto proyectual en una Especialización Docente en Didáctica de las Ciencias de la Computación, *Revista Teyet* no. 27, pp. 9-19, 2020.
- [8] N. Colussi & P. Viale. "Actividades de Programación Grupales para Primer año de la Licenciatura en Ciencias de la Computación - Experiencias Didácticas en el Aula". In *Jornadas Argentinas de Didáctica de la Programación (JADiPro)*, 2019.
- [9] N. Colussi, P. Viale, N. Monjelat. "Vidriera de proyectos: una modalidad de evaluación posible en tiempos de virtualidad". En *II Workshop de Innovación y Transformación Educativa. "Transformación Digital. Desafíos de la Educación Superior"*. 2021
- [10] Viale P., Deco C., Bender C. Enseñando Pensamiento Computacional en el Ciclo Básico de las carreras de Ingeniería. *50° Jornadas Argentinas de Informática*. pp 42-51. SAEI, Simposio Argentino de Educación en Informática. CABA. 2021.
- [11] Viale P., Deco C., Bender C. Introduciendo conocimientos sobre el Pensamiento Computacional en los primeros años de las carreras de Ingeniería. pp 921-926 *COINI 2020: XIII Congreso Internacional de Ingeniería Industrial*. CABA. edUTecNe, 2021
- [12] García Angarita M., Deco C., Bender C., Collazos C. "Una Propuesta para el Desarrollo de Pensamiento Computacional en Niños y Jóvenes". *Revista Teyet*. Nro. 30. pp 16-27. Diciembre 2021. La Plata, Argentina.

DESARROLLO Y EVALUACIÓN DE COMPETENCIAS EN LA INGENIERÍA DE SOFTWARE EN UN ENTORNO VIRTUAL DE APRENDIZAJE COLABORATIVO

Nicolás Battaglia, Carlos Neil, Marcelo De Vincenzi, Charles Maldonado,
Facundo Oliva, Jorge Lomoro, Fabian Maffei

Universidad Abierta Interamericana. Facultad de Tecnología Informática.
Centro de Altos Estudios en Tecnología Informática. Buenos Aires. Argentina
{Nicolas.Battaglia, Carlos.Neil, MeDevincenzi}@uai.edu.ar,
{CharlesRoy.MaldonadoDuarte, JuanFacundo.Oliva, Jorge.Lomoro,
Fabian.Maffei}@alumnos.uai.edu.ar

RESUMEN

Dos años después del recordado “año de la pandemia”, la sociedad naturalizó muchos de los cambios que fueron obligados en ese contexto. En particular, las instituciones educativas han adoptado la educación híbrida (en la mayoría de los casos) y la educación a distancia en otros. Por este motivo, han tenido que enfrentar diversos desafíos tales como reinventar sus modelos pedagógicos para que el proceso de enseñanza aprendizaje continúe de manera ininterrumpida y capacitar su cuerpo docente. Todo esto no hubiese sido posible sin los cambios constantes de las TIC. Cambios que nos hicieron reflexionar sobre la efectividad del método tradicional de enseñanza: los graduados deben ser competentes, deben conocer los saberes de un área de conocimiento y aplicarlos en futuros contextos profesionales. Un doble desafío: diseñar una infraestructura para poder brindar servicios de educación mediada por la tecnología y que los estudiantes puedan aprender en forma colaborativa para acreditar sus competencias. En este proyecto se propone la integración de plataformas tecnológicas para el aprendizaje colaborativo y evaluación basada en competencias, en cursos vinculados con la ingeniería de Software.

Palabras clave: Aprendizaje Colaborativo, Aprendizaje Basado en Competencias, Ingeniería de Software, Trabajo Colaborativo, UML.

CONTEXTO

El proyecto de investigación se implementa en la facultad de Tecnología Informática de la Universidad Abierta Interamericana (UAI). Estudiantes de la universidad de 2° a 5° año de la carrera de Ingeniería en Sistemas Informáticos, trabajan de manera articulada en las siguientes asignaturas relacionadas: 1) Metodologías de

Desarrollo de Sistemas I y II (2° año). 2) Bases de Datos, Trabajo de Diploma y Trabajo de Campo I (3° año). 3) Seminario de Aplicación Profesional y Trabajo Final de Ingeniería (5° año).

El proyecto, por otro lado, está en línea con el documento “Aportes para la implementación de un modelo de formación orientado a competencias” [1], donde se formulan las bases para el rediseño del plan de estudios centrado en competencias de la terminal Ingeniería en Sistemas de Información / Informática.

1. INTRODUCCIÓN

Como se pudo observar durante los años 2020 y 2021, uno de los aspectos que más sufrió el impacto del COVID-19, sin duda, fue la educación. Las instituciones educativas tuvieron que reestructurarse en muy poco tiempo para poder continuar operando de manera virtual y así lograr, parcialmente, sus objetivos. Por medio de las TIC, gran parte de los estudiantes continuaron estudiando, pero de manera virtual. Esto presentó nuevos desafíos: formar a los docentes en esta nueva modalidad y, además, implantar y adaptar sistemas pedagógicos e informáticos que permitieron lograr estos objetivos, en general, de manera parcial. Si bien esta realidad aceleró el proceso de cambio, también es cierto que durante los últimos años la educación ha ido cambiando sus métodos de aprendizaje centrado en el docente a métodos de aprendizaje centrados en el estudiante [2].

Desde el comienzo del proyecto de investigación [3], [4] afirmamos que el cambio y la evolución generó la necesidad de repensar el sistema educativo. Esta afirmación y el COVID-19 como catalizador, nos obliga a ahondar, aún más, en las implicancias de la tecnología en la educación.

Uno de los cambios más relevantes de los últimos tiempos es la capacidad actual de poder acceder a información de manera ubicua. Esto

está vinculado, sin dudas, con la evolución en las TIC [5].

De esta manera, gracias a Internet y a los dispositivos móviles, el estudiante puede aprender de manera virtual, utilizar herramientas sincrónicas (videoconferencias, chat, etc.) y asincrónicas (foros y redes que promueven la interacción con otras personas), facilitándoles el aprendizaje colaborativo. Por este motivo, el trabajo colaborativo, las TIC y la enseñanza se pueden integrar para crear entornos de aprendizaje colaborativo asistido por computadora. Esta integración se basa en el impacto social del trabajo colaborativo y las TIC en la educación que transforma la manera tradicional de enseñanza/aprendizaje. Esto conlleva varios cambios, entre ellos el Aprendizaje Basado en Competencias (ABC) y centrada en el estudiante y el impacto de las TIC en los procesos de enseñanza y aprendizaje, para lo cual es necesario también modificar y adaptar los criterios y estrategias de evaluación motivados por esta disrupción [6].

Por un lado, el trabajo colaborativo apoyado en el concepto de Zona de Desarrollo Próximo (ZDP) propuesto por Vygotski a principios del siglo XX [7] y, por el otro, las TIC en el marco de las propuestas de *e-learning* o *blended learning*. Esto sumado al desarrollo creciente de las TIC junto con el concepto de trabajo colaborativo, conforman los entornos de Trabajo Colaborativo Asistido por Computadora (CSCW). Este concepto, integrado en entornos de enseñanza y aprendizaje colaborativo, dio origen a los entornos de Aprendizaje Colaborativos Asistidos por Computadora (CSCL) [8]. En el ámbito académico, los nuevos estándares para la acreditación de carreras de ingeniería en la República Argentina [1], [9] que proponen el ABC y están centrado en el estudiante, dan cuenta de la necesidad de la integración de ambos modelos, además de considerar todas las herramientas de seguimiento y evaluación que sean requeridas por este paradigma. Por otro lado, el estudio teórico de la Ingeniería de Software (IS) no es suficiente para comprender y resolver los problemas de cooperación y colaboración que surgen durante el desarrollo de un proyecto informático [10]. Los estudiantes suelen centrar su esfuerzo en aspectos técnicos y asumen que los inconvenientes vinculados al trabajo en equipo no impactarán en el proyecto. En virtud de resolver el problema planteado previamente, Daniele et al. [11] proponen, para mejorar los aspectos comunicacionales, la integración de las plataformas CSCL con entornos especializados en la resolución de problemas prácticos y

técnicos en un proyecto de desarrollo de software; además, esta propuesta potencia el aprendizaje de la IS y permite el desarrollo de competencias profesionales y sistemáticas en el desarrollo de software [4]. De la misma manera, tal como sucede en los procesos tradicionales de enseñanza, en los entornos CSCL, la evaluación cumple un rol muy importante. Según el estudio realizado en [12], en el modelo de enseñanza y aprendizaje de la IS mediado por tecnología, los estudiantes cumplen un papel de mayor protagonismo y responsabilidad y, en consecuencia, los docentes asumen un nuevo rol cuando es el estudiante quien adquiere mayor autonomía en su proceso de aprendizaje a través de entornos virtuales. Esta autonomía es una característica distintiva de la enseñanza centrada en el estudiante. En este contexto, la evaluación es entendida como un proceso que promueve el aprendizaje con una finalidad formativa, más que como un proceso de control de resultados. Por consiguiente, un proceso específico de enseñanza y aprendizaje en un entorno virtual colaborativo requiere también de un proceso de evaluación, autoevaluación y coevaluación (evaluación 360°) y seguimiento acorde al modelo de aprendizaje colaborativo planteado por el concepto CSCL [6]. El énfasis en la educación centrada en el estudiante permite retomar el modelo de ABC, que también promueve reformular y adaptar los criterios y estrategias de evaluación motivados por la disrupción tecnológica que da origen a los entornos CSCL.

En resumen, en este proyecto nos enfocamos en la enseñanza y el aprendizaje del modelado de software utilizando el estándar UML como lenguaje para especificar, construir, visualizar y documentar sistemas informáticos. Para esto proponemos el rediseño de un entorno CSCL específico para dicha área de conocimiento, con un enfoque centrado en el ABC. Otro aspecto importante es el énfasis en la utilización de rúbricas analíticas para la evaluación de competencias.

2. APRENDIZAJE BASADO EN COMPETENCIAS

La universidad debe preparar al estudiante en un entorno de competencias genéricas y específicas que les permitan afrontar los problemas que plantea esta sociedad cambiante. McClelland [13], en la búsqueda de una alternativa a las pruebas de aptitud e inteligencia tradicionales, desarrolló el concepto de "competencia" definido como una característica subyacente de una persona que le permite demostrar un desempeño superior en un

determinado rol o situación, haciendo la diferencia entre personas con desempeño excelente versus personas con desempeño promedio. Por otro lado, el concepto de competencia engloba todo un conjunto de conocimientos, procedimientos y actitudes que se combinan, coordinan e integran, para que el individuo logre «saber conocer», «saber hacer», «saber ser» [14]. El dominio de estos saberes le permite ser «capaz de» actuar con eficacia y eficiencia en situaciones que se le presenten [15]. En este sentido, la competencia no reside en los recursos (capacidades) sino en cómo éstos se movilizan para lograr un objetivo. Además, el aprendizaje por competencias ha promovido toda una renovación de las teorías psicopedagógicas de los aprendizajes, dando lugar a la evolución de los esquemas de referencia de la formación de profesionales [15].

3. EVALUACIÓN DE COMPETENCIAS

Tal como planteó Perrenoud [16], las competencias son invisibles y sólo son abordables a través de los desempeños observables, por lo tanto, en línea con Zapata [17] éstas deben ser observables y evaluables y esto sucede, básicamente, por medio de las conductas que los sujetos generan ante diversos problemas en contextos cambiantes. Esto plantea la necesidad de la sustitución o coexistencia de procedimientos, herramientas y formas de evaluar los saberes por medio de procesos descriptivos o conceptuales (exámenes, pruebas de destreza, problemas teóricos, etc.) por otros métodos donde se tenga en cuenta la simulación real de casos prácticos, con metodologías de indagación, de investigación formativa, con elaboración de proyectos, etc.

Como consecuencia, el ABC requiere un sistema de evaluación variado, debido a que cada competencia tiene componentes muy distintos que necesitan procedimientos diversos para ser evaluados correctamente [18]. Si bien existen diversas metodologías de evaluación de competencias [19], en la actualidad uno de los mejores instrumentos, son las rúbricas, debido a su facilidad de uso y a la oportunidad de obtener aspectos complejos, imprecisos y subjetivos [12].

Según Zapata [20], cuando se evalúa en un entorno virtual de aprendizaje, es necesario también la comunicación de los resultados que ayudan al estudiante a comprender el nivel de adquisición de una competencia determinada. En estos entornos, es fundamental el concepto tradicional de retroalimentación o *feedback* propuesto por los entornos virtuales. Sin

embargo, en los entornos virtuales colaborativos, es necesario medir también el nivel de participación de un estudiante dentro de un equipo de trabajo, por ejemplo, por medio de la cantidad de mensajes enviados, tomando en cuenta aspectos semánticos y de relevancia en cuanto a cómo son estos mensajes, su influencia en otros estudiantes, si son respondidos, etc.

4. RÚBRICAS ANALÍTICAS

La necesidad de instrumentar la evaluación para que facilite la mejora del aprendizaje implica, primero, plantearla como una tarea de aprendizaje; segundo, involucrar a los estudiantes y, finalmente, ofrecer sus resultados como retroalimentación [21]. Estos tres objetivos pueden abordarse mediante el uso de rúbricas que, como instrumento de evaluación, permiten valorar los aspectos complejos, imprecisos y subjetivos de un tema/problema, aportando una evaluación fácilmente interpretable, justa y transparente para profesores y estudiantes [22].

Las rúbricas no sólo evalúan, también enseñan; la autoevaluación basada en su uso contiene los criterios de evaluación que el estudiante ha de utilizar para valorar su trabajo [23]. Como instrumento de evaluación, brinda un conjunto de criterios de calidad relacionados con las competencias a evaluar y, mediante descriptores y diferentes niveles de logro, ponen de manifiesto no sólo el incremento cuantitativo de los estudiantes, sino también el salto cualitativo; demuestran cuánto han aprendido y lo bien que lo han hecho [24]. La rúbrica no es solamente una herramienta de evaluación sino, también, una estrategia de aprendizaje; en el proceso de autoevaluación, el estudiante no resuelve un problema y luego se autoevalúa, lo hace con las consideraciones establecidas en los criterios de evaluación que le permite reflexionar sobre el mismo y actuar en consecuencia. Por otro lado, su utilización se potencia en los contextos virtuales de enseñanza y aprendizaje, ya que su diseño y uso empoderan al docente, al estudiante y a la institución promoviendo la cultura de la evaluación auténtica en la educación [25].

Por otro lado, la evaluación no debe ser un proceso distinto del aprendizaje, es una oportunidad para promoverlo [26]. La evaluación 360° tiene como objetivo desarrollar un modelo que cubra todos los aspectos relevantes de la evaluación de un proceso a partir de la retroalimentación, permitiendo elevar la calidad y la mejora continua tanto del desempeño docente como del estudiante, es decir, de la enseñanza y del aprendizaje [27]. En este sentido,

las rúbricas favorecen la autoevaluación debido a que poseen dos características fundamentales: los criterios de evaluación y los estándares de calidad [23].

5. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El presente trabajo está identificado como proyecto de investigación y desarrollo en el Centro de Altos Estudios en Tecnología Informática (CAETI), dentro de la línea de investigación Sociedad del Conocimiento y Tecnologías aplicadas a la educación. Además, el proyecto se ramifica en diferentes áreas temáticas específicas, a saber: 1) Analítica de aprendizaje: entendemos este concepto como “la medición, recopilación, análisis e informe de datos sobre los estudiantes y sus contextos, con el fin de comprender y optimizar el aprendizaje y los entornos en los que se produce”. 2) Estilos de aprendizaje y personalización de la enseñanza: el aprendizaje personalizado es un enfoque educativo cuya finalidad es que éste el aprendizaje se ajuste a las fortalezas, necesidades, habilidades e intereses de cada estudiante. La identificación de estilos de aprendizaje permite adaptar las estrategias de enseñanza a las características del estudiante. El objetivo es identificar, para cada estudiante, su estilo de aprendizaje para personalizar las estrategias de enseñanza. 3) Estudios empíricos sobre modelos de enseñanza aprendizaje: la herramienta UAI Case [28] se enfoca en la enseñanza y el aprendizaje del modelado de software utilizando el estándar UML. Existen muchos enfoques de enseñanza aprendizaje, desde los tradicionales hasta los más innovadores. El objetivo es realizar estudios empíricos que permitan establecer la ganancia de aprendizaje cuando se utiliza UAI Case respecto de otras estrategias de enseñanza. 4) Inteligencia artificial aplicada a la educación: el aprendizaje en entornos virtuales permite al estudiante organizar sus tiempos y no depender de un encuentro sincrónico para cumplir los objetivos, sino que puede hacerlo de manera desconectada, en cualquier momento y lugar. Sin embargo, el estudiante necesita obtener *feedback* del profesor y para eso depende de sus tiempos y espacios para poder responder consultas y corregir actividades. En base a esto, consideramos que la incorporación de herramientas cognitivas de educación, permitirán asistir a los estudiantes en tiempo real durante el proceso de aprendizaje y de esta manera permitir una mayor independencia del estudiante para poder realizar sus actividades. 5) Gamificación: la Ingeniería de

software es una actividad centrada en las personas [29], que no solo incluye aspectos de tipo técnico o tecnológico, sino que además está fuertemente influenciado por las competencias blandas que poseen los miembros del equipo. Bajo esta perspectiva, el proceso de desarrollo de software enfrenta desafíos relacionados con el tratamiento de factores sociales y humanos que influyen en el desempeño de los equipos de trabajo y su repercusión en el éxito o no de los proyectos de desarrollo de software [30]. La investigación se centrará en la adecuación de estrategias didácticas basadas en la gamificación [31], que se puedan diseñar y construir en un entorno virtual colaborativo, para la formación en ingeniería de software [32]. Utilizando las estrategias didácticas mediadas por recursos TIC, que apunten a mejorar la enseñanza, aprendizaje y evaluación de competencias específicas de la ingeniería de software. 6) Mecanismo estandarizado de propuesta de mejoras en las herramientas de enseñanza virtual en Plataformas educativas: dado que las plataformas de educación virtual son un recurso utilizado en la enseñanza superior y universitaria, es menester tener un estándar de calidad en su uso e implementación [34]. Existen plataformas comerciales que ya vienen preparadas para ser multifuncionales [35] pero estas deberían ser útiles para el desarrollo de las actividades de enseñanza y aprendizaje [36]. El objetivo es crear un mecanismo estandarizado de propuesta de mejoras en las herramientas de enseñanza virtual ubicua (específicamente, en plataformas educativas) para obtener una perspectiva general de los aspectos básicos que debe tener una plataforma para satisfacer los requisitos de la educación virtual de calidad [37].

6. RESULTADOS OBTENIDOS/ESPERADOS

i. Resultados obtenidos:

Se han obtenido los siguientes resultados: 1) El desarrollo, diseño e implementación del prototipo de una herramienta CASE colaborativa ubicua multiplataforma, denominada UAI Case [28]. 2) El diseño de una plataforma académica colaborativa multiplataforma para evaluación, seguimiento, interacción y coordinación de proyectos informáticos. 3) La definición y especificación de los bloques funcionales necesarios para determinar un proceso específico en la enseñanza y aprendizaje de modelados en la IS. 4) El diseño de un modelo de rúbrica analítica implementada en un entorno Web.

ii. Resultados Esperados:

Se esperan los siguientes resultados: 1) Diseñar herramientas que permitan integrar el concepto de ABC junto con herramientas que faciliten evaluar y realizar seguimientos en entornos colaborativos (UAI Case). 2) Generar indicadores en tiempo real que permitan medir la evolución de los saberes y competencias en el ABC. 3) Obtener un conjunto de indicadores que permitan medir el rendimiento de un equipo virtual de trabajo durante el proceso de enseñanza y aprendizaje de la IS. 4) Realizar la evaluación empírica del prototipo de la herramienta UAI Case por medio de la definición de un método de evaluación basado en métricas. 5) Desarrollar sistemas de tutorías inteligentes mediante la implementación de tutores cognitivos que imiten el rol del profesor, guíen el desarrollo del aprendizaje y ofrezcan pistas a los estudiantes cuando están atascados en un problema. 6) Utilizar analíticas del aprendizaje que permitan, mediante la medición, recopilación y análisis de datos de los estudiantes durante el proceso de enseñanza aprendizaje, detectar los errores más comunes y proporcionen una respuesta en tiempo real al estudiante. 7) Personalizar las estrategias de enseñanza a partir de los estilos de aprendizaje de los estudiantes.

7. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está formado, por 14 estudiantes de posgrado (doctorado en ciencias informáticas, maestría en tecnología informática y especialización en ingeniería de software) quienes están desarrollando sus respectivas tesis y trabajos finales en temas vinculados al proyecto: Fernando Parra, Jorge Lomoro, Silvia Poncio Gabriela Iannantuoni, Jorge Zarate, Marcelo Monferrato, Juan Facundo Oliva, Charles Maldonado, Nelson Garrido, Alejandro Downar, Mariano Bucher, Nicolás Repetti, Fabian Maffei y Darío Kiryczun.

8. BIBLIOGRAFÍA

- [1] C. Neil, N. Sotomayor, R. Muñoz, P. Cristaldo, B. Parra de Gallo, and J. C. Calloni, *Aportes para la implementación de un modelo de formación orientado a competencias*, Ira. Ed. Ciudad Autónoma de Buenos Aires: Universidad Abierta Interamericana, 2021.
- [2] A. Schiter, N. Battaglia, and C. Neil, "El Porqué de las Rúbricas para la Evaluación y la Autoevaluación," *Congr. Nac. Ing. Informática - Sist. Inf. CONAIISI*, 2018.
- [3] C. Neil, M. De Vincenzi, N. Battaglia, and R. Martínez, "Herramientas Colaborativas Multiplataforma en la Enseñanza de la Ingeniería de Software," *XVIII Work. Investig. en Ciencias la Comput. (WICC 2016, Entre Ríos, Argentina)*, 2016.
- [4] N. Battaglia, C. Neil, M. De Vincenzi, and J. P. Beltramino, "Competency-based learning in collaborative virtual platforms," *Int. Inst. Innov. Technol. 10th IIITEC Int. Symp. Innov. Technol. ISIT2019. Cuzco, Perú.*, 2019.
- [5] G. Lavigne, M. Ovando, J. Sandoval, and L. M. Salas, "Exploración preliminar del aprendizaje colaborativo dentro de un entorno virtual," *Actual. Investig. en Educ.*, 2012.
- [6] N. Battaglia, "Integración de una Herramienta CASE en un Entorno Académico Colaborativo para la Enseñanza de Ingeniería de Software," Universidad Abierta Interamericana, 2017.
- [7] L. Vygotsky, "The collected works of LS Vygotsky: Problems of the theory and history of psychology," 1997.
- [8] G. Stahl, T. Koschmann, and D. Suthers, "Computer-supported collaborative learning: An historical perspective," 2006.
- [9] CONFEDI, *Propuesta de estándares de segunda generación para la acreditación de carreras de ingeniería en la República Argentina "Libro Rojo de CONFEDI"*. 2018.
- [10] P. Bouillon, J. Krinke, and S. Lukosch, "Software engineering projects in distant teaching," 2005, doi: 10.1109/CSEET.2005.31.
- [11] M. Daniele, M. Uva, P. Martelloto, and G. Picco, "Aplicación de herramientas CASE a la enseñanza de Ingeniería de Software: Gestión de la Configuración de Software y Testing Funcional," 2010.
- [12] N. Battaglia, R. Martínez, M. Otero, C. Neil, and M. De Vincenzi, "Autoevaluación Colaborativa por medio de Rubricas en Entornos Virtuales de Enseñanza y Aprendizaje," 2016.
- [13] D. C. McClelland, "Testing for competence rather than for 'intelligence'.,", *Am. Psychol.*, 1973, doi: 10.1037/h0034092.
- [14] S. Tobón, "La Formación Basada en Competencias en la Educación Superior: El enfoque complejo.", 2009.
- [15] J. T. Fernández and C. R. Bueno, "Evaluación de competencias profesionales en educación superior: Retos e implicaciones," *Educ. XXI*, 2016, doi: 10.5944/educXXI.12175.
- [16] P. Perrenaud, *Construir competencias desde la escuela*. JC Sáez, 2008.
- [17] A. Fernández March, "La evaluación orientada al aprendizaje en un modelo de

- formación por competencias en la educación universitaria,” *REDU. Rev. Docencia Univ.*, 2011, doi: 10.4995/redu.2010.6216.
- [18] A. Villa Sánchez and M. Poblete Ruiz, *Aprendizaje Basado En Competencias Una Propuesta Para La Evaluación*. 2007.
- [19] J. Cubero-Ibáñez, M. S. Ibarra-Sáiz, and G. Rodríguez-Gómez, “Propuesta metodológica de evaluación para evaluar competencias a través de tareas complejas en entornos virtuales de aprendizaje,” *Rev. Investig. Educ.*, vol. 36, no. 1, pp. 159–184, 2018, doi: 10.6018/rie.36.1.278301.
- [20] M. Zapata, “Evaluación de competencias en entornos virtuales de aprendizaje y docencia universitaria,” *Revista de Educación a Distancia. Sección de Docencia Universitaria en la Sociedad del Conocimiento.*, vol. 1, no. 1DU, 2010.
- [21] I. Á. Valdivia, “Evaluación del aprendizaje en la universidad: Una mirada retrospectiva y prospectiva desde la divulgación científica,” *Electron. J. Res. Educ. Psychol.*, vol. 6, no. 14, pp. 235–272, 2008.
- [22] M. García Hípola, “Diseño de la auto, co-evaluación y rúbrica como estrategias para mejorar el aprendizaje,” *VI Jornadas sobre Innovación Docente en Arquít.*, 2018, doi: 10.5821/jida.2018.5433.
- [23] E. Panadero, J. A. Tapia, and J. A. Huertas, “Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education,” *Learn. Individ. Differ.*, vol. 22, no. 6, pp. 806–813, 2012, doi: 10.1016/j.lindif.2012.04.007.
- [24] M. P. G. Sanz, “La evaluación de competencias en Educación Superior mediante rúbricas: un caso práctico,” *Rev. electrónica Interuniv. Form. del Profr.*, vol. 17, no. 1, p. 106, 2014.
- [25] N. Battaglia, C. Neil, and M. De Vincenzi, “Software Engineering Competence-Based Learning in Collaborative Virtual Environments,” *V IEEE Congr. Mund. Educ. en Ing.*, 2021.
- [26] V. Torres-Sanz, J. Tramullas, J. A. Sanguesa, P. Garrido, and F. J. Martínez, “Rúbricas como estrategia de evaluación en entornos TICS [Rubrics as an evaluation strategy in ICT environments],” 2017.
- [27] Y. Jiménez Galán, M. González, and J. Hernández, “Modelo 360 ° para la evaluación por competencias (enseñanza-aprendizaje),” *Innovación Educ. IPN*, vol. 10, no. 53, pp. 43–53, 2010.
- [28] N. Battaglia, C. Neil, M. De Vincenzi, R. Martínez, and J. P. Beltramino, “UAI case: desarrollo y evaluación de competencias en la ingeniería de software en un entorno virtual de aprendizaje colaborativo,” *XXI Work. Investig. en Ciencias la Comput. (WICC 2019, Univ. Nac. San Juan).*, 2019.
- [29] L. María, M. Suárez, V. Daniel, G. Vera, and M. Suarez, “Gamificación en la enseñanza de Técnicas de Elicitación de Requisitos,” *researchgate.net*, 2021.
- [30] L. Machuca-Villegas, G. P. Gasca-Hurtado, L. M. R. Tamayo, and S. M. Puente, “Elementos de gamificación en el contexto de ingeniería de software,” *RISTI - Rev. Iber. Sist. e Tecnol. Inf.*, vol. 2020, no. E27, pp. 718–732, 2020.
- [31] M. Souza, R. Moreira, and E. Figueiredo, “Playing the Project: Incorporating Gamification into Project-based Approaches for Software Engineering Education,” in *sol.sbc.org.br*, 2020, pp. 71–80, doi: 10.5753/wei.2019.6618.
- [32] M. Del Carmen Pegalajar Palomino, “Implications of gamification in Higher Education: A systematic review of student perception,” *Rev. Investig. Educ.*, vol. 39, no. 1, pp. 169–188, 2021, doi: 10.6018/RIE.419481.
- [33] C. Maldonado and D. Pungutá, “Estrategias didácticas apoyadas en TIC para el desarrollo de competencias transversales,” *Educ@cion en Context.*, vol. 7, no. 13, pp. 228–253, 2021.
- [34] J. D. Britos *et al.*, “Los MOOC como propuesta para la estandarización de la calidad educativa,” 2016.
- [35] J. Onrubia, “Aprender y enseñar en entornos virtuales: actividad conjunta, ayuda pedagógica y construcción del conocimiento,” *Rev. Educ. a Distancia*, no. 50, 2016, doi: 10.6018/red/50/3.
- [36] F. Molina, N. Da, S. Ignacio, A.-C. Juan, and J. Goyeneche, “Evaluación y monitoreo de plataformas educativas,” *colibri.udelar.edu.uy*, 2021.
- [37] I. Aguaded, R. M.-S.-R. R. iberoamericana de, and undefined 2015, “Criterios de calidad para la valoración y gestión de MOOC,” *redalyc.org*, vol. 18, no. 2, pp. 119–143, 2015.

Diseño Participativo para desarrollar Recursos Educativos para la Enseñanza de las Ciencias de la Computación en la Escuela Secundaria. Una Máquina de Turing en la Escuela

Daniel Dolz Gerardo Parra Jorge Rodríguez
{ddolz, gparra,j.rodrig}@fi.uncoma.edu.ar

Grupo de Investigación en Lenguajes e Inteligencia Artificial
Departamento de Teoría de la Computación - Facultad de Informática
UNIVERSIDAD NACIONAL DEL COMAHUE

Resumen

La presencia de las Ciencias de la Computación en el currículum escolar está captando la atención en todo el mundo. Las iniciativas tienen como objetivo que la población estudiantil de todos los niveles tenga acceso a los conceptos centrales de la disciplina.

Las tendencias curriculares actuales promueven desarrollar un recorrido amplio por las áreas de conocimiento. Entre ellos, los relacionados al área de Teoría de la Computación.

Los recursos desenchufados, sobre todo los desarrollados en el ámbito de la iniciativa *CS Unplugged*, están ampliamente difundidos. Sin embargo no existen suficientes para cubrir conceptos sobre Teorías de la Computación.

Esta iniciativa se enmarca dentro de la Línea destinada a desarrollar recursos didácticos y evaluar su efectividad. En particular se trabaja sobre la construcción y evaluación de una colección de recursos orientados a facilitar la enseñanza de conceptos relacionados a las Máquinas de Turing.

Se trabaja en el contexto de los enfoques metodológicos basados en la investigación y

diseño participativos definidos por esta Línea de Investigación y Desarrollo.

Palabras Clave: Educación En Ciencias De La Computación, Escuela Secundaria, Recursos Educativos Desenchufados, Participatory Design, Máquina de Turing.

Contexto

Esta línea de investigación y desarrollo se enmarca, por un lado, en el contexto de los temas de interés promovidos por el Grupo de Investigación en Lenguajes e Inteligencia Artificial (GILIA), de la Facultad de Informática de la Universidad Nacional del Comahue (UNCo).

En particular, se enmarca en el ámbito del proyecto de investigación de la Facultad de Informática denominado *Modelos Formales, Agentes Inteligentes y Aplicaciones para la Enseñanza de las Ciencias de la Computación*, presentado en la Convocatoria 2022.

Por otro lado, el trabajo se desarrolla teniendo en cuenta el Convenio Marco de Colaboración firmado durante 2016 entre la Facultad de Informática y el Ministerio de Educación de la Provincia del Neuquén. Particularmente, se trabaja con el Consejo Provincial de Educación

de la Provincia de Neuquén.

1. Introducción

La presencia de contenidos de Ciencias de la Computación en el currículum escolar está captando la atención en todo el mundo. Las iniciativas tienen como objetivo que la población estudiantil de todos los niveles educativos, tenga acceso a los conceptos centrales de la disciplina [8,9,10].

Las tendencias curriculares actuales promueven desarrollar un recorrido amplio por las áreas de conocimiento. Entre ellos, los relacionados al área de Teoría de la Computación.

Si bien en la República Argentina se observa un proceso dispar e iniciativas con cierta preponderancia del área de Algoritmos y Programación, se nota que progresivamente se tiende a un recorrido más amplio.

En esta dirección, la provincia de Neuquén aprobó recientemente el diseño curricular para la escuela secundaria que incorpora, entre otros, contenidos sobre Teoría de la Computación [11].

Este tipo de conocimiento es descripto como abstracto y complejo de aprender para estudiantes sin formación previa en la disciplina. Los recursos desenchufados, sobre todo los desarrollados en el ámbito de la iniciativa *CS Unplugged*, están ampliamente difundidos y suelen ofrecer un primer contacto satisfactorio [1, 2, 3].

Aunque inicialmente fueron pensados para actividades no escolarizadas, sobre todo para eventos de divulgación científica, las escuelas los adoptan como forma de ofrecer primeros contactos con conceptos abstractos sobre

computación.

Sin embargo, aún no existe suficiente investigación sobre su efectividad en el ámbito escolar o acerca de la forma que deben adoptar para adecuarse a estos contextos institucionalizados.

Por otra parte, si bien existe un abanico amplio de tópicos disciplinares cubiertos por este tipo de recursos, existen otros que están presentes en las propuestas curriculares no cubiertos por estas colecciones de recursos. Esta es la situación de un conjunto de conceptos fundamentales sobre Teoría de la Computación, como los relacionados a las Máquinas de Turing [4].

Este trabajo se enmarca dentro de la Línea de Investigación y Desarrollo destinada a producir recursos didácticos para enseñar Ciencias de la Computación y evaluar su efectividad en el ámbito de la educación secundaria.

En particular, plantea trabajar sobre el desarrollo y evaluación de una colección de recursos educativos desenchufados orientados a facilitar la enseñanza de conceptos relacionados a las Máquinas de Turing y a nociones introductorias de computabilidad [4,5].

En este contexto se trabaja en el marco de los enfoques metodológicos basados en la investigación y el diseño participativos definidos específicamente por esta Línea de Investigación y Desarrollo [12].

El resto del artículo está organizado de la siguiente manera. A continuación, describimos la línea de investigación y desarrollo. En la sección 3, presentamos los resultados obtenidos y esperados. Finalmente, en la

sección 4, comentamos aspectos relacionados a la formación de recursos humanos en el marco de esta línea, así como del proyecto de investigación en su conjunto.

2. Línea de investigación y desarrollo

El trabajo presentado en este artículo se enmarca en la Línea de Investigación y Desarrollo denominada *Recursos Educativos Desenchufados para la Enseñanza de las Ciencias de la Computación en la Escuela Secundaria*.

La colección de recursos educativos adopta la forma de juegos de mesa donde se debe seguir una serie de pasos para lograr llegar a un resultado. Las reglas del juego están definidas por los mecanismos de funcionamiento de las Máquinas de Turing.

Los siguientes aspectos conforman la caracterización de los grupos destinatarios de la propuesta:

- Año de estudio: primeros años de la escuela secundaria.
- Conocimientos disciplinares asumidos: Ninguno
- Tiempo destinado: aproximadamente 40 minutos.
- Tamaño del grupo recomendado: se recomienda que los equipos estén conformados por 3 a 4 integrantes. De una actividad participan varios equipos.

Aspectos metodológicos:

A continuación, se presentan algunas opciones metodológicas adoptadas para este trabajo.

- Promover la colaboración, la utilización de recursos físicos actúa como facilitador de la actividad grupal.
- Distribuir la complejidad, cada persona asume la responsabilidad de una pieza de la máquina. La complejidad conceptual se distribuye en el grupo.
- Aprender jugando, se traslada la mecánica de juego al ámbito del aprendizaje para lograr mejores resultados en términos disciplinares y de desarrollo de habilidades blandas.
- Aprender de la experiencia, centrado en producir conocimiento abstracto y conceptual a partir de reflexionar sobre experiencias concretas.

Diseño participativo.

El proceso de diseño participativo [12], definido para este trabajo se inicia convocando al equipo de trabajo afectado al diseño de una primera versión para el Recurso Educativo destinado a enseñar conceptos fundamentales sobre Máquinas de Turing.

Este equipo se compone de investigadores, estudiantes universitarios y docentes de escuelas secundarias, que en conjunto exploran el área de conocimiento, determinan enlaces curriculares para la actividad y entrecruzan el conocimiento del que disponen sobre la enseñanza de la computación en el ámbito de la escuela secundaria.

En segunda instancia, se amplía el equipo de diseño convocando a un grupo mayor de profesores de informática de escuelas secundarias. El grupo ampliado, explora el recurso elaborado aportando nuevas perspectivas acerca del mismo. En conjunto se revisa el material producido y se elabora un reporte que informa sobre las características valoradas positivamente y acerca de aquellas

que requieren ajustes.

Una vez aplicados los ajustes, la colección de recursos estará a disposición de los docentes, en formato de Recurso Educativo Abierto, para ser utilizados en sus aulas. Estos ciclos de adecuación progresiva anteceden a los estudios destinados a evaluar la efectividad de la colección de recursos.

Estudiar la efectividad de la tipología particular de recursos.

Desde la perspectiva metodológica, esta línea se ubica en el ámbito de la Investigación Acción Participativa [6,7]. La participación comunitaria se expresa, en este caso, en la acción desplegada por docentes de escuelas secundarias.

Como primera aproximación, la colección de producciones elaboradas en el marco del diseño participativo asume el lugar de conjeturas de carácter teórico y práctico en relación a las posibilidades didácticas de este tipo de recurso educativo.

Estas elaboraciones disponen de cierto grado de flexibilidad que las hace susceptibles a ser ajustadas para situarse en contextos educativos singulares sin perder rigurosidad disciplinar. El equipo de investigación junto al grupo de docentes que participa del estudio trabajan para acomodar el material a las particularidades de la escuela secundaria.

El equipo conformado por investigadores y docentes realiza el diseño experimental destinado a estudiar la efectividad de este tipo de recursos educativos.

Como cierre de este proceso, se analizan las experiencias desarrolladas en las escuelas

como forma de contribuir a la mejora de la producción de conocimiento teórico.

3. Resultados obtenidos

En el contexto de las actividades desarrolladas en esta línea se logró definir un enfoque metodológico basado en el Diseño Participativo para orientar procesos de elaboración de recursos educativos destinados a enseñar Ciencias de la Computación.

En la misma línea, se definió un modelo en el contexto de la Investigación Acción Participativa destinado a estudiar la efectividad de este tipo de recurso didáctico y aportar elementos que contribuyan al refinamiento de las construcciones teóricas en desarrollo.

Se realizó un trabajo de campo de carácter experimental con intención de evaluar si estos mecanismos eran accesibles al docente y con qué facilidad ellos consideraban que eran trasladables al aula.

Se tomó el recurso Human Area Network para abordar la enseñanza del tema redes de computadoras, específicamente, se evaluó la efectividad como dispositivo para introducir el tópico Medios Físicos de Transmisión de Datos [13].

El diseño y estudio de efectividad fueron orientados por los modelos teóricos definidos en ésta Línea de Investigación.

4. Formación de Recursos Humanos

Se espera que el desarrollo de esta línea de Investigación contribuya a la formación de recursos humanos en el campo de las Ciencias

de la Computación en la Educación. En este sentido, dos integrantes del proyecto están finalizando la Maestría en Ciencias de la Computación de la Facultad de Informática, UNCo. Además, otro integrante del equipo está finalizando la Maestría en Enseñanza en Escenarios Digitales, que desarrollan de manera conjunta las Universidades Nacionales de Cuyo, Comahue, Patagonia Austral, Patagonia San Juan Bosco, San Luis, Chilecito y La Pampa.

Un integrante docente del proyecto de investigación está finalizando su tesis de Licenciatura en Ciencias de la Computación en la temática del grupo. Los integrantes alumnos del proyecto también desarrollan sus tesis en temas relacionados.

Por otra parte, actualmente se están desarrollando al menos cinco tesis de Licenciatura en temas de interés del grupo de investigación.

5. Referencias

- [1] T. Bell and J. Vahrenhold. Cs unplugged—how is it used, and does it work? In *Adventures Between Lower Bounds and Higher Altitudes*, pages 497–521. Springer, 2018.
- [2] T. Nishida, S. Kanemune, Y. Idosaka, M. Namiki, T. Bell, and Y. Kuno. A cs-unplugged design pattern. *ACM SIGCSE Bulletin*, 41(1):231–235, 2009.
- [3] R. Taub, M. Armoni, and M. Ben-Ari. Cs unplugged and middle-school students’ views, attitudes, and intentions regarding cs. *ACM Transactions on Computing Education (TOCE)*, 12(2):1–29, 2012.
- [4] H. Lewis and C. Papadimitriou. *Elements of the Theory of Computation*. Second Edition. Prentice Hall, 1998.
- [5] J. Hopcroft, R. Motwani and J. Ullman. *Introduction to Automata Theory, Languages and Computation (3rd Edition)*. Addison Wesley, 2006.
- [6] A. M. Colmenares E. Investigación-acción participativa: una metodología integradora del conocimiento y la acción. *Voces y Silencios. Revista Latinoamericana de Educación*, 3(1):102–115, 2012.
- [7] J. Martí. *La investigación-acción participativa: estructura y fases*. 2017.
- [8] K-12 Computer Science Framework Steering Committee. *The K–12 Computer Science Framework*. ACM, 2016.
- [9] F. Sadosky. CC – 2016 Una propuesta para refundar la enseñanza de la computación en las escuelas argentinas. Fundación Sadosky, Argentina, 2013.
- [10] R. Society. *After the reboot: Computing education in UK schools*. Policy Report, 2017.
- [11] C. P. de Educación de la Provincia de Neuquén. *Diseño Curricular Jurisdiccional de los tres primeros años de la Escuela Secundaria Neuquina*. Resolución N° 1463/18, 2018.
- [12] B. DiSalvo, J. Yip, E. Bonsignore, and D. Carl. *Participatory design for learning*. In *Participatory design for learning*, pages 3–6. Routledge, 2017.
- [13] Dolz, Daniel, et al. *Recursos Educativos Desenchufados para la Enseñanza de las Ciencias de la Computación en la Escuela Secundaria*. XV Congreso Nacional de Tecnología en Educación y Educación en Tecnología (TE&ET 2020). Neuquén, 2020.

Ciencias de la Computación en la Escuela Secundaria

Percepciones Sociales y Formación Docente Inicial

Jorge Rodríguez ¹

Marcos Manuel Cortez ²

Sandra Boari ³

j.rodri@fi.uncoma.edu.ar mmcortez@neuquen.gov.ar sandraboari@gmail.com

¹*Grupo de Investigación en Lenguajes e Inteligencia Artificial*
Departamento de Teoría de la Computación - Facultad de Informática
UNIVERSIDAD NACIONAL DEL COMAHUE

²*Dirección Provincial de Educación Secundaria*

³*Dirección Provincial de Educación Superior*
Consejo Provincial de Educación

MINISTERIO DE GOBIERNO Y EDUCACIÓN DE LA PROVINCIA DE NEUQUÉN

Resumen

Las Ciencias de la Computación han intensificado su lugar de relevancia en los diseños curriculares para la educación obligatoria en diversos países del mundo.

En este contexto, se plantea que la Formación Docente Inicial constituye una pieza clave para incidir positivamente en el desarrollo y consolidación de los procesos de inclusión en todos los niveles de la educación obligatoria de manera sostenible y rigurosa.

Por otra parte las percepciones que las partes interesadas tienen respecto a los roles, propósitos y posición de la computación en la escuela secundaria influyen en las formas que efectivamente adopta la computación en las aulas.

Este trabajo propone realizar una revisión sistemática que describa rigurosamente cuáles paradigmas y enfoques curriculares se expresan en los documentos curriculares para la formación docente inicial destinada a cubrir los espacios curriculares del área informática de la educación secundaria.

Por otra parte, se plantea realizar un estudio centrado en describir las percepciones de la comunidad educativa acerca del lugar de la computación en las propuestas curriculares, como también establecer puntos de continuidad y

ruptura con iniciativas curriculares desarrolladas en el país.

Palabras Clave: CIENCIAS DE LA COMPUTACIÓN, CURRÍCULUM, ESCUELA SECUNDARIA, ENSEÑANZA DE LA COMPUTACIÓN, FORMACIÓN DOCENTE INICIAL.

Contexto

Esta propuesta se ubica en el contexto de las iniciativas promovidas por el Grupo de Investigación en Lenguajes e Inteligencia Artificial de la Facultad de Informática y del Convenio Marco de Colaboración firmado durante 2016 entre la Facultad de Informática y el Ministerio de Educación de la Provincia del Neuquén.

El convenio tiene como principal objetivo contribuir recíprocamente al desarrollo de actividades de investigación, formación de recursos humanos y promoción de la enseñanza de las Ciencias de la Computación durante la escolaridad obligatoria. Este trabajo se desarrolla en el ámbito del proyecto de investigación *Modelos Formales, Agentes Inteligentes y Aplicaciones para la Enseñanza de las Ciencias de la Computación* que está financiado por la Univer-

sidad Nacional del Comahue a través de la Secretaría de Ciencia y Técnica. El proyecto tiene prevista una duración de cuatro años a partir de enero del 2022.

1. Introducción

Desde comienzos del siglo XXI las Ciencias de la Computación han intensificado un recorrido que las ubica en un lugar relevante en los diseños curriculares para la educación obligatoria en diversos países del mundo [10, 11]. Este proceso cobra impulso a partir de los reportes “*Running On Empty: The Failure to Teach K–12 Computer Science in the Digital Age*”, en Estados Unidos, y *Shut down or restart?: The way forward for computing in UK schools*, en Reino Unido [14, 18].

Asimismo, en la última década, se observa que en Argentina la incorporación de las Ciencias de la Computación en los niveles obligatorios del sistema educativo se presenta como un horizonte que se aproxima de manera constante. El desarrollo de propuestas en esta dirección, por parte del Consejo Federal de Educación, la iniciativa Program.ar, un conjunto de Universidades Nacionales y los gobiernos provinciales [3, 4, 5] requieren ampliar la presencia en las agendas de las políticas educativas, ya que la computación aún está poco representada en los diseños curriculares de la mayoría de las jurisdicciones [12].

En este contexto, se plantea que la Formación Docente Inicial y Continua constituye una pieza clave para incidir positivamente en el desarrollo y consolidación de los procesos de inclusión de las Ciencias de la Computación en los niveles de la educación obligatoria de manera sostenible y rigurosa [7, 14, 19].

Por otra parte las percepciones sobre los roles, propósitos y posición que las partes interesadas tienen respecto de la computación en la escuela secundaria influyen fuertemente en las formas que efectivamente adopta la computación en las aulas [6, 9, 15].

En este contexto, la investigación disponible en este campo de estudio en la órbita nacional es insuficiente, aun cuando existe consenso

acerca de que la Formación Docente Inicial y las percepciones que las comunidades educativas elaboran sobre el lugar de las Ciencias de la Computación en la educación secundaria, describiendo categorías centrales para comprender la situación actual y para la definición de políticas públicas.

Esto se observa en la carente información sistematizada en relación a la oferta de Formación Docente Inicial para el área informática, sus características, los paradigmas y enfoques a los que adhieren y el grado de articulación con las políticas públicas para con la enseñanza de las Ciencias de la Computación en la educación obligatoria. Pocos estudios investigan sobre la opinión de la comunidad educativa respecto al cuerpo de conocimiento que debe integrar la propuesta formativa, el momento e intensidad con que deben ser incorporados, como también el sentido de su inclusión.

El contexto descripto expone la necesidad de dar continuidad a la Línea de Investigación destinada específicamente a desarrollar estudios, referidos a aspectos curriculares de la Educación en Ciencias de la Computación, que contribuyan a describir y comprender la situación de la enseñanza de la disciplina en el ámbito de la educación secundaria en el país [2, 17, 16].

Se propone realizar una revisión sistemática que describa rigurosamente cuáles paradigmas y enfoques curriculares se expresan en los documentos curriculares para la formación docente inicial destinada a cubrir los espacios curriculares del área informática de la educación secundaria.

Por otra parte, esta Línea de Investigación plantea realizar un estudio centrado en describir las percepciones de la comunidad educativa acerca del lugar de la computación en las propuestas curriculares, como también establecer puntos de continuidad y ruptura con iniciativas curriculares desarrolladas en el país.

El resto del artículo está organizado de la siguiente manera. A continuación, se presenta la línea de investigación. En la sección 3, se describen los resultados iniciales obtenidos y los que se espera obtener producto del desarrollo de esta investigación. Finalmente, en la sección 4,

se comentan aspectos relacionados a la formación de recursos humanos en el marco de esta línea, así como del proyecto de investigación en su conjunto.

2. Línea de investigación

Para el desarrollo de esta línea se propone trabajar en forma articulada sobre los siguientes aspectos:

- *Revisión sistemática de los planes de estudios de las carreras de profesorado destinadas a la formación inicial de docentes de informática.*

Se plantea realizar una revisión sistemática y situada en la Educación Superior a nivel nacional, específicamente en las carreras terciarias y universitarias de Formación Docente Inicial en Informática o en Ciencias de la Computación.

Se procura 1. identificar los paradigmas para la selección de conocimientos computacionales presentes en los Planes de Estudios, 2. aportar una síntesis que permita comprender los paradigmas y enfoques en los que se expresan las Ciencias de la Computación, 3. realizar aportes que colaboren con el análisis de la situación actual y la definición de perspectivas curriculares futuras y 4. evidenciar la presencia de las áreas de conocimiento de las Ciencias de la Computación.

El insumo primario para desarrollar el estudio será del tipo documental para lo cual se recopilarán los documentos curriculares de cada una de las carreras de Formación Docente Inicial en Informática y Ciencias de la Computación pertenecientes a Instituciones Terciarias y Universitarias, tanto estatales como privadas con el propósito de identificar, evaluar e interpretar la información de interés para la investigación. La búsqueda y extracción se realizará siguiendo criterios que se definirán explícitamente de acuerdo al procedimiento propuesto por Kitchenham y Booth [1, 8].

- *Percepciones de la comunidad educativa acerca del lugar de las Ciencias de la Computación en la escuela secundaria neuquina.*

Se busca una mejor comprensión del lugar que estudiantes, docentes, padres, madres y equipos de conducción escolar asignan a la computación en el trayecto de la educación secundaria. El estudio se enfoca sobre tres aspectos: 1. Paradigmas, Roles y Propósitos de la computación en la escuela secundaria, 2. Posición o año académico de la computación en el plan de estudio y 3. Enfoques para la selección de contenidos.

Se busca conectar las percepciones o representaciones que la comunidad educativa elabora acerca de la correspondencia existente, entre las Ciencias de la Computación y la educación secundaria con los constructos discutidos en el contexto de las investigaciones desarrolladas en este campo de conocimiento. Por otra parte, se explora el tipo de enlace que existe entre éstas representaciones y las iniciativas curriculares que promueven la ampliación de la presencia de las Ciencias de la Computación en los diseños curriculares para la educación del nivel secundario.

El proceso de recolección de datos se sostendrá sobre dos dispositivos: entrevistas semi estructuradas y encuestas, que de conjunto permitirán obtener la información para avanzar en el estudio. A partir de las mismas se indagará a estudiantes, padres, madres, docentes e integrantes de equipos de conducción de un conjunto de escuelas secundarias que constituyan una muestra representativa de las instituciones educativas de la jurisdicción.

Por consiguiente se pretende averiguar la incidencia de las iniciativas curriculares en las representaciones, analizar posibles puntos de encuentro entre los lugares que la comunidad educativa asigna a la computación y los enfoques curriculares vigentes.

Los estudios presentados en esta sección buscan aportar elementos que permitan refinar la definición de los paradigmas y enfoques formulados en esta Línea de Investigación. Tienen la intención de contribuir a la constitución de categorías de análisis satisfactorias para explicar la situación de la computación en los diseños curriculares.

La relevancia de estos estudios radican en que permiten mejorar la comprensión del sentido pedagógico con el que la computación participa de las propuestas de enseñanza oficiales para la educación secundaria.

3. Resultados obtenidos y esperados

En el contexto de las actividades desarrolladas en ésta línea se identificaron paradigmas y enfoques curriculares que permiten describir con rigurosidad la situación de la computación en las propuestas curriculares destinadas a la escuela secundaria [2, 16].

Se desarrollaron revisiones sistemáticas tendientes a describir La posición de las Ciencias de la Computación en el Diseño Curricular para la Escuela Secundaria Argentina [12, 13].

Se analizaron diseños curriculares para la escuela secundaria correspondientes a 21 jurisdicciones del país. El principal resultado obtenido consiste en identificar que las Ciencias de la Computación están subrepresentadas en las propuestas formativas para la educación secundaria en Argentina, el enfoque curricular predominante es el integrado y la incorporación de la computación se produce con un alto grado de dispersión en relación a los enfoques y perspectivas propuestos.

En los estudios que dan continuidad a la Línea de Investigación se esperan obtener los siguientes resultados:

- Presentar una síntesis que contribuya a describir y comprender la situación de la Formación Docente Inicial para el área de computación en Argentina.

- Aportar elementos que ayuden a identificar Propósitos, Roles, Paradigmas, Enfoques y Posición que la comunidad educativa percibe para la computación en la educación secundaria.
- Aportar información que permita refinar la definición de Enfoques y Paradigmas como categorías de análisis satisfactorias para explicar la situación de la computación en los diseños curriculares.

En esta sección se presentaron los resultados obtenidos y esperados que se ubican como una continuidad de la Línea de Investigación destinada a estudiar aspectos curriculares relacionados a la incorporación de la computación en la educación obligatoria.

4. Formación de Recursos Humanos

Se espera que el desarrollo de esta línea de Investigación contribuya a la formación de recursos humanos en el campo de las Ciencias de la Computación en la Educación.

En este sentido dos de los autores de este artículo cursan maestrías orientadas a conocer, comprender y analizar procesos relacionados con las tecnologías en la educación. Ambos se encuentran en proceso de elaboración de tesis en temáticas vinculadas a esta Línea de Investigación.

Referencias

- [1] A. Booth, A. Sutton, and D. Papaioannou. *Systematic approaches to a successful literature review*. Sage, 2016.
- [2] G. Branchini, M. M. Cortez, M. d. I. Á. Pedemonte, and J. Rodríguez. Las ciencias de la computación en el currículum escolar. In *XXI Workshop de Investigadores en Ciencias de la Computación (WICC 2019, Universidad Nacional de San Juan)*, 2019.

- [3] Consejo Federal de Educación. Res 263/15. *Resoluciones CFE*, 2015.
- [4] Consejo Federal de Educación. Res 343/18. *Resoluciones CFE*, 2018.
- [5] Consejo Provincial de Educación. Res 1463/18. *Resoluciones CPE*, 2018.
- [6] G. . Gallup. Images of computer science: Perceptions among students, parents, and educators in the us, 2015.
- [7] N. Granor, L. A. DeLyser, and K. Wang. Teals: Teacher professional development using industry volunteers. In *Proceedings of the 47th acm technical symposium on computing science education*, pages 60–65. ACM, 2016.
- [8] B. Kitchenham and S. Charters. Guidelines for performing systematic literature reviews in software engineering. EBSE technical report. *Keele University & Department of Computer Science University of Durham*, 2007.
- [9] K. Kori and P. Luik. Upper-and lower-secondary students’ motivation to study computer science. In *International Conference on Informatics in Schools: Situation, Evolution, and Perspectives*, pages 69–78. Springer, 2020.
- [10] O. McGarr and K. Johnston. Curricular responses to computer science provision in schools: current provision and alternative possibilities. *The Curriculum Journal*, 31(4):745–756, 2020.
- [11] G. Ottestad and G. B. Gudmundsdottir. Information and communication technology policy in primary and secondary education in europe. *Second handbook of information technology in primary and secondary education*, pages 1–21, 2018.
- [12] J. Rodríguez and M. Cortez. La posición de las ciencias de la computación en el diseño curricular para la escuela secundaria argentina: Una revisión sistemática. *Electronic Journal of SADIO (EJS)*, 19(2):136–150, 2020.
- [13] J. Rodríguez, M. Cortez, and S. Boari. Explorando el lugar de las áreas de conocimiento de las ciencias de la computación en la escuela secundaria argentina: Una revisión sistemática. *50JAIIO - SAEI*, 2021.
- [14] Royal Society UK. *Shut down or restart?: The way forward for computing in UK schools*. Royal Society, 2012.
- [15] M. Säde, R. Suviste, P. Luik, E. Tõnisson, and M. Lepp. Factors that influence students’ motivation and perception of studying computer science. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, pages 873–878, 2019.
- [16] S. Sommer, M. E. Cornejo, M. Cortez, and J. Rodríguez. El lugar de las ciencias de la computación en el currículum de la escuela secundaria argentina. In *XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018, Universidad Nacional del Nordeste)*, 2018.
- [17] S. Sommer, M. E. Cornejo, J. Rodríguez, and L. Cecchi. Aproximando las ciencias de la computación a la escuela secundaria. In *XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, ITBA, Buenos Aires)*, 2017.
- [18] C. Wilson, L. A. Sudol, C. Stephenson, and M. Stehlik. *Running on empty: The failure to teach k–12 computer science in the digital age*. ACM - CSTA, 2010.
- [19] H. S. Yadav Arman, Sarah Gretter. Expanding computer science education in schools: understanding teacher experiences and challenges. *COMPUTER SCIENCE EDUCATION*, page 2, 2016.

INTELIGENCIA ARTIFICIAL: ESTRATEGIAS APLICADAS A PROBLEMAS EMERGENTES

Laura Cecilia Díaz Dávila, Silvia Edith Arias, José Daniel Britos, Guillermo Eduardo Zaballo, Adolfo Vignoli, Gladys Moyano, Leonor Vega, Gisela Hirschfeld, Néstor Martiarena, Sandro Comerci, Lucas Signorini, Matías Almarcha, Matías Nicolás Sosa, Trinidad Fernandez Tucci, Dana Del Bono, Facundo Nieto, Cristian Ezequiel Morilla, Giuliana Mangini, Victoria Carro, Analía Gonzales Wernli.

Departamento de Computación
Facultad de Ciencias Exactas, Físicas y Naturales
Universidad Nacional de Córdoba
Av. Vélez Sarsfield 1611, Córdoba
0351-4334409

laura.diaz@unc.edu.ar, sil_var@hotmail.com, dbritos@unc.edu.ar, guillermo.zaballo@unc.edu.ar,
adolfo.vignoli@unc.edu.ar, gladys.moyano.566@unc.edu.ar, leonor.teresa.vega@unc.edu.ar,
gise320@gmail.com, nestor.ricardo.martiarena@gmail.com, comercisandro@gmail.com,
lucassignorini257@gmail.com, matias.almarcha@mi.unc.edu.ar,
mnsosa@mi.unc.edu.ar, trinidad.fernandez.tucci@mi.unc.edu.ar, danadbon@gmail.com,
facundonieto@mi.unc.edu.ar, cristian.morilla@mi.unc.edu.ar, giuliana.mangini@mi.unc.edu.ar,
victoriacarrouba@gmail.com, analiawernli@hotmail.com

RESUMEN

El objetivo de esta presentación es compartir experiencias del equipo Aprendizaje Inteligente en su proceso de transición desde el proyecto de investigación actual hacia el que dará comienzo en 2023.

En este trabajo se muestran resultados del proyecto vigente: “Un Enfoque Integral Para Propiciar Cursos Abiertos On Line Desde La Universidad Nacional De Córdoba”. Con un enfoque multidisciplinario, desarrolló acciones de formación de investigadores en tecnologías novedosas como: procesos de digitalización, blockchain, sistemas, software, plataformas educativas y herramientas del campo de la Inteligencia Artificial; para coadyuvar a los diferentes enfoques requeridos en el contexto tecnológico actual caracterizado por sus vertiginosos cambios y el aprendizaje colaborativo. Profundizó en el conocimiento del modelo del estudiante, en sus características socioeconómicas, académicas y cognitivas, mediante el uso de Tecnologías Inteligentes de Explotación de la Información (TIEI), no sólo para mejorar los procesos de aprendizaje y evaluación, sino además descubrir patrones de comportamiento relevantes para procesos decisionales en la gestión de Educación Superior (ES). Aprendizaje Inteligente orienta sus intereses a

la construcción de un prototipo MOOC <https://aprendizajeinteligente.fcefyn.unc.edu.ar/>, cuyo objeto de aprendizaje se asocia a disciplinas académicas de ciencia, tecnología, ingeniería y matemáticas (CTIM - STEM science, technology, engineering and mathematics). La investigación aplicada que aquí se comparte pretende favorecer acceso, continuidad y permanencia, mitigando el desgranamiento. Convocó a un equipo consolidado, cuyos antecedentes con diferentes matices datan de más de una década. También se incorporan en esta presentación, los emergentes que justifican las reorientaciones de las líneas de investigación y las acciones desarrolladas que contribuyen a definir la arquitectura del nuevo proyecto que dará inicio en 2023.

Palabras clave: Cursos abiertos online, Tecnologías Inteligentes, Enfoque multidisciplinario

CONTEXTO

En la categoría de “Proyecto Consolidar III”, este equipo contó con el acompañamiento de dos “Proyectos Formar” de la Facultad de Ciencias Exactas, Físicas y Naturales y un proyecto Consolidar de la Facultad de Ciencias de la Comunicación, en el marco del Programa

“Apropiación Social del Conocimiento y la Tecnología”.

Cabe resaltar que en 2021 despedimos a dos de los investigadores principales de Aprendizaje Inteligente: Laura Mónica Vargas (Doctora en Ciencias de la Ingeniería con una sólida formación matemática) y José Luis Gallopo (Ingeniero, Coordinador del Ciclo de articulación de Matemática en la FCEfyN). Esta irreparable pérdida nos obliga a trabajar para alcanzar los logros que aún están pendientes, a repensar nuestros objetivos en función de las fortalezas del equipo y el contexto actual, y a ofrecer un homenaje en su memoria.

Actualmente, el equipo de Aprendizaje Inteligente continúa trabajando para finalizar el plan de trabajo actual y rediseñar sus nuevas orientaciones que ponen énfasis en un enfoque multidisciplinario de la Inteligencia Artificial para dar solución a problemas emergentes que impactan significativamente en la mejora de calidad de vida de las personas; en el marco de las recomendaciones de la UNESCO de noviembre de 2021[1] en la conferencia 41 de Naciones Unidas a la que adhirieron 68 países miembros.

Los proyectos de investigación involucrados son acreditados y financiados por la Secretaría de Ciencia y Técnica de la UNC.

INTRODUCCIÓN

Los objetivos, el plan de trabajo y los resultados esperados del proyecto actual se pueden consultar en memorias de WICC de 2019.[2] Así mismo, cabe aclarar que la mayoría de fuentes externas en las que basa esta investigación, son citadas en [2], por lo cual no se referencian en éste trabajo. Se citan las contribuciones realizadas por el equipo durante el periodo en cuestión.

En Argentina, el acceso a las Universidades públicas y el desgranamiento asociado a la masividad despierta el interés por ofrecer propuestas creativas acordes a esta realidad. La cual cobra singular relevancia en el contexto sanitario mundial que puso en evidencia las necesidades estructurales desatendidas en tal sentido, y además aceleró procesos asociados a

las nuevas tecnologías de información y comunicación, exigiendo un mayor esfuerzo para trabajar en la mejora de las condiciones de la apropiación social de estas tecnologías; a la vez que ofrece una oportunidad histórica de acelerar este proceso. Esto en tiempos en que la construcción de saberes constituye un desafío importante.

Las características propias de Latinoamérica devienen en una importante responsabilidad social hacia los estudiantes.[2] Además se imponen los MOOC de las grandes corporaciones Universitarias. En este escenario, Aprendizaje Inteligente avanzó hacia la construcción de propuestas creativas para atender a la articulación entre niveles medio y universitario, por una parte al Ciclo General de Conocimientos Básicos (CGCB) por otra, específicamente en Matemática para Ingeniería en la Universidad Nacional de Córdoba (UNC). Abordó un tratamiento integral, aunque no completo, atendiendo a las políticas de la gestión de gobierno de la UNC, los paradigmas de Construcción Colaborativa de Aprendizajes y las Nuevas Estrategias de Evaluación y Acreditación de Saberes.

En primer lugar, [2] la investigación se orientó a descubrir patrones de comportamiento del estudiante, principal actor del complejo escenario, a partir del uso de Inteligencia Artificial y sus disciplinas soporte,[3] para enriquecer las múltiples aplicaciones que emergen de este conocimiento. [2]

Esta línea lleva inmanentes dos objetivos fundamentales: favorecer la adquisición de competencias (saber SER, saber CONOCER y saber HACER) de los investigadores en formación que forman parte de Aprendizaje Inteligente y además, concebir el uso de herramientas de Inteligencia Artificial como aspecto necesario para transitar el camino hacia el diseño del MOOC.

En segundo lugar, [2] se focalizó en la mejora del aprendizaje en Matemática para las carreras de la FCEfyN en sus procesos de articulación entre los niveles medio y universitario como área temática prioritaria escogida para la intervención, justificada en: las experticias del equipo, las demandas

detectadas en los CGCB y en los Consorcios InterUniversidades y la emergencia de la pandemia.

En el contexto actual de aceleración del uso de plataformas digitales, las más tradicionales como Manhattan, Claroline, Moodle y Sakai, parecen no adecuarse a los nuevos requerimientos en cuanto a la masividad y la interactividad con los videos y otros recursos.

Aprendizaje Inteligente tuvo la experiencia de enfrentar los problemas asociados a estos cursos que demandan equipos multidisciplinarios, recursos físicos importantes y además una inversión significativa en formación de esta matriz productiva de MOOC. En el apartado de resultados se relatan los avances materializados en un prototipo desplegado en enero de 2021 para el CINEU de Matemática de la FCEFyN, su estado actual y los resultados esperados

En tercer lugar, [2] muy ligado al anterior, desde este espacio se hicieron contribuciones hacia la Ingeniería de producto de MOOC para favorecer la sistematización de esta nueva industria, similar a la otrora explosiva del cine, que nace en el campo de la Educación.

Los resultados describen brevemente el enfoque utilizado que, nuevamente privilegió formar a los investigadores más jóvenes y posibilitar su invaluable contribución, a partir del desarrollo de un trabajo final de carrera de una de las integrantes para alcanzar el grado de Ingeniera Industrial, dirigida por dos de los Investigadores con responsabilidad formal. Este desarrollo no estaba contemplado en el plan de trabajo del proyecto, sin embargo el equipo consideró que resulta de una gran contribución a la situación actual que tampoco era concebible en aquél momento. Por otra parte, no dificultó la ejecución del plan de trabajo sino más bien que vino a suplir algunas de las actividades previstas que, por razones ajenas, no pudieron llevarse a cabo tales como los cursos de para la Confederación Mundial de Actividades SubAcuáticas (CMAS) a partir del convenio suscripto con la FAAS en 2017. En esta línea también se abordó el desarrollo de experticias en los procesos de producción, implementación y evaluación de guiones

audiovisuales, una **sublínea de investigación**.

En este sentido, se desarrolló un Curso de Postgrado “Plataforma Open EDX-LMS Aplicaciones y Usos” aprobado por Resolución HCD: 1475/2019 y que, de alguna manera se constituyó en el espacio aglutinante de todas las áreas de investigación durante el 2020, en situación de mayor crisis de pandemia.

En cuarto lugar, [2] basado en las evidencias de las plataformas que brindan servicios a través de Internet requieren conexiones seguras en sus transacciones y en el avance vertiginoso de las aplicaciones de blockchain, en 2018 este equipo inició su investigación aplicada en transacciones entre instituciones educativas y en voto electrónico en línea. Así surgieron dos proyectos finales de carrera, ambos dirigidos y desarrollados por integrantes de Aprendizaje Inteligente, que culminaron entre 2020 y 2021, dando espacio a la reformulación de nuevos horizontes que se comparten en el apartado de resultados.

En suma, estas son las áreas que guiaron nuestras acciones en este momento histórico tan singular de amenazas y oportunidades emergentes. Además, desde aquí se conciben los nuevos horizontes de tecnologías aplicadas, manteniendo la fuerte convicción de: el enfoque multidisciplinario, la formación de diversos actores y las acciones orientadas al bien común; en el marco de las recomendaciones internacionales para contribuir a la apropiación responsable de las tecnologías disruptivas.

1. RESULTADOS DE LAS LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

1. Investigación en aspectos relativos a la plataforma Open edX.

Curso de postgrado con transferencia efectiva ya referido.[4][5]

Trabajo Final de carrera,[6] cuyo objetivo es contribuir con una aproximación a la ingeniería de estos cursos concebidos como producto.[7] Aborda el desarrollo del ciclo de vida de un MOOC y aspectos inherentes al proceso de producción, con un enfoque

integral para favorecer su diseño, desarrollo y despliegue.[8][9]

Se inspira en el curso que emergió en el seno del equipo de investigación. El producto estuvo disponible a inicios de 2021 en su versión beta, en una plataforma de Open edX y actualmente se incorporan mejoras para desplegarlo desde el Campus Virtual de la UNC como un curso de interés general para los diversos actores de la articulación entre nivel medio y universitario. [10]

2. Producción de un prototipo de MOOC en Open edX para Matemática de CINEU de la FCEFyN. Es el prototipo al que se refiere el anterior, sus resultados fueron publicados. [11][12]

3. Descubrimiento de patrones de comportamiento del estudiante de ES, aplicando IA. Publicados en [13] en TE&ET 2020 [14] , CADI 2018[15] , RIAT 2019[16] , Guadalajara 2021[5]. Los requerimientos asociados al conjunto de datos consideran el desempeño académico al momento de ingresar a la universidad, la continuidad de sus estudios y aspectos del entorno socioeconómico de la persona. Se contaron con más de dos mil setecientos registros de alumnos ingresantes 2016 y 2017 de la FCEFyN. Como asignatura objetivo o clase se tomó Matemática de CINEU. Basado en el diseño de Explotación de la Información con herramientas de IA aplicado a estudiantes de las carreras de Ingeniería de la FCEFyN - UNC y en la de estudiantes de ciencias económicas [17].

4. Investigación aplicada en blockchain: exploración de plataformas y aplicación a voto electrónico en línea y a transacciones entre instituciones educativas.[18] Los dos equipos publicaron sus avances en RIAT 2019 y el informe completo de cada trabajo final de grado está disponible en la biblioteca de la FCEFyN, ambos de la Escuela de Ingeniería en Computación.[19]

5. Acciones para un enfoque integral: Se desarrolló una estrategia de gestión de redes sociales, incluyendo la creación de la página institucional de Aprendizaje Inteligente cuya arquitectura incluye, por ejemplo, los enlaces a las publicaciones del equipo y otros hitos relevantes, se diseñó y llevó a cabo un proceso

para indagar sobre preferencias de recursos audiovisuales cuyos resultados se tuvieron en cuenta en el desarrollo del curso para matemática y publicados en eventos a tal fin. [20][21]

6. Formación de grado y posgrado: Transversal a los demás, para cada uno de ellos, los resultados materializan cursos con transferencia efectiva, desde la asignatura de IA integrándose al proyecto, trabajos finales de carrera, oportunidades de desarrollar roles de auxiliares de docencia como Mentores en la Diplomatura de Ciencia de Datos de la FAMAFA, de prácticas profesionales supervisadas en el Laboratorio de Innovación e Inteligencia Artificial de la Universidad de Buenos Aires (UBA-IALAB), de obtención de becas para el Programa Multidisciplinario en Inteligencia Artificial Aplicada, también de UBA-IALAB, formación en gestión de redes sociales, sólo por mencionar algunas de las más relevantes que significaron una retroalimentación efectiva para los objetivos del proyecto. Es posible concluir que se ha conformado una estructura académica-científica y de transferencia efectiva, una relación virtuosa cuya finalidad es contribuir con la aplicación de estas tecnologías hacia el bien común, formando jóvenes responsables con habilidades integrales que constituyen el orgullo de nuestro equipo.

2. NUEVOS EJES TEMÁTICOS

En razón del carácter integral del abordaje del proyecto y de las consideraciones expresadas al cierre del apartado anterior, se esbozan a continuación las líneas temáticas que ya emergen para el siguiente periodo, incorporándose una fuerte presencia de las recomendaciones de ética para la IA que la UNESCO publicó a fines de 2021 y que ya fueron referidas. Merecen ser distinguidas:

1. La primera línea da continuidad a la de IA aplicada a la Educación. Desde allí emerge poniendo énfasis en la posibilidad de escalar las arquitecturas y los modelos de IA diseñados y desarrollados, aplicables a diversos problemas de diferentes dominios que tengan

como finalidad el bien común, más allá del ámbito de la ES, tales como las administraciones públicas y los espacios del Estado, independientemente de su ámbito y jurisdicción. En tal sentido ya se desarrollaron ensayos de arquitecturas, basadas en modelos de IA para imágenes, en la detección de patrones visuales en textos. Esta lógica de solución de respuestas en múltiples ámbitos de problemas asociados al bien común.

La insuficiente digitalización de documentos públicos en latinoamérica, como los expedientes judiciales y las historias clínicas en hospitales, demandan soluciones que pueden escalar a partir de ensayos de laboratorio como los que se han desarrollado. Estas acciones han tenido singular repercusión en varios actores del ecosistema nacional de la IA y se encuentra en elaboración el primer documento para su publicación. Es importante destacar los acuerdos de colaboración que permiten profundizar en esta dirección provenientes de UBA-IALAB.

2. Aprendizaje Inteligente incluye líneas de vanguardia en su plan de trabajo como una estrategia necesaria en su ámbito de interés. Un subequipo se dedica específicamente a la investigación exploratoria de: el concepto de Metaverso, las diversas aplicaciones del juego, los ambientes inmersivos y la computación cuántica. El objetivo consiste en definir líneas de investigación aplicada para el próximo periodo.

3. Favorecer el desarrollo de los investigadores en formación del equipo en las diferentes áreas disciplinares, en el marco de las recomendaciones de UNESCO para la ética de la IA aplicada.

3. FORMACIÓN DE RECURSOS HUMANOS

La configuración del equipo es variable. Actualmente lo lideran Directora, Co directora y cuatro Investigadores Responsables. Finalizaron tres tesinas de grado, hay en curso: cuatro prácticas profesionales supervisadas, dos tesinas de grado y dos estudiantes de Doctorado. Lo integran veinte investigadores en los roles de: Director, Responsable, En Formación o Colaborador.

4. BIBLIOGRAFÍA

- [1]UNESCO. 2021. "Recomendación sobre la ética de la Inteligencia Artificial". 41ª sesión de la Conferencia General.
- [2]L. C. Díaz Dávila et al. 2019. "Un enfoque integral para propiciar cursos abiertos on line desde la Universidad Nacional de Córdoba". XXI Workshop de Investigadores en Ciencias de la Computación.
- [3]Díaz Dávila, Comerci, Arias- 2020. "Tecnologías Disruptivas: Inteligencia Artificial aplicable a la Gestión de Políticas Públicas en Educación Superior en contextos de masividad". Libro Red de Investigadores de Apropiación de la Tecnología (RIAT) 2020.
- [4]Díaz Dávila, Britos, Hirschfeld. 2020. "La Pandemia. Acciones para facilitar el aprendizaje en Matemática durante el ingreso a carreras de ingeniería". TE&ET 2020.
- [5]Díaz Dávila, Britos, Arias. 2020. "MOOC en matemática para el acceso a la educación superior- Avances hacia la educación 4.0. un enfoque integral" XXVIII Encuentro Internacional de Educación a Distancia. U.de G. Méjico
- [6]Fernandez Tucci T., Britos D., Diaz Dávila L. 2021. "Ingeniería de Producto para un Prototipo de MOOC". CAEDI 2021.
- [7]Meléndez, et al. 2016. Guía práctica: gestión, producción, infraestructura y control de calidad para MOOC. Nuevas Ideas en Informática Educativa. Volumen 12, p. 372 - 377. Santiago de Chile
- [8]Lladós, et al. 2018. Guía práctica para hacer un MOOC. Universidad Nacional de Córdoba.
- [9]Soberanis, et al. 2019. Implicaciones de la industria 4.0 en la educación superior.
- [10]Mooc-Maker. 2018. Libro guía sobre aspectos académicos de gestión en relación a los MOOCs en la Educación Superior.
- [11]Galoppo, Zaballo, Arias, Diaz Dávila. 2021. "Percepciones de los estudiantes sobre materiales audiovisuales para el aprendizaje de matemática en el ingreso a las carreras de Ingeniería". EMCI 2021
- [12]Galoppo, Vega, Moyano. 2020. MOOC EN MATEMÁTICA PARA EL ACCESO A LA EDUCACIÓN SUPERIOR. XXVIII Encuentro Internacional de Educación a Distancia. U.de

G. Méjico

[13]Díaz Dávila, Comerci Sandro, 2019.

“Contribuciones de Cátedra”. Primeras Jornadas de Inteligencia Artificial del Litoral

[14]Díaz Dávila L., Britos D., Hirschfeld G.

2020. La Pandemia. Acciones para facilitar el aprendizaje en Matemática durante el ingreso a carreras de Ingeniería. TE&ET 2020

[15]Díaz, L. 2018. “Conocimiento Inteligente para los Procesos Decisionales en Carreras de Ingeniería”. CAEDI 2018

[16]Díaz, L., Comerci, S. 2019. “Conocimiento inteligente del estudiante universitario en contextos de masividad”. RIAT 2019

[17]Díaz L., Martins S., Garcia-Martinez R. 2015. “Descubrimiento de Patrones Socioeconómicos de Población Estudiantil de Carreras de Ingeniería Basado en Tecnologías de Explotación de Información” TE&ET 2015.

[18]Giralda, Vilches, Britos, Díaz Dávila. 2020. Sistema de voto electrónico en línea. Mejoras con blockchain. Trabajo final de carrera. Biblioteca FCEFyN.

[19]Giralda, Vilches, Britos, Diaz Dávila. 2019. “Avances hacia un prototipo de Sistema de Votación Electrónico con base de datos distribuida - Tecnología BlockChain”. Libro RIAT 2020.

[20]Martirena, Diaz Dávila, Arias. 2020. “MOOC EN MATEMÁTICA PARA EL ACCESO A LA EDUCACIÓN SUPERIOR. INDAGANDO EL COMPORTAMIENTO Y LAS PERCEPCIONES DEL ESTUDIANTE”. XXVIII Encuentro Internacional de Educación a Distancia. U.de G. Méjico.

[21]Hirschfeld, Zaballos, Gonzalez Guerni. 2020. “MOOC EN MATEMÁTICA PARA EL ACCESO A LA EDUCACIÓN SUPERIOR- RETOS DE LOS RECURSOS AUDIOVISUALES” XXVIII Encuentro Internacional de Educación a Distancia. U.de G. Méjico.

IS – Ingeniería en Software

Avances en la incorporación de estrategias innovadoras en los procesos de enseñanza y de aprendizajes de la programación

Gustavo ASTUDILLO¹, Silvia BAST¹, Leandro CASTRO¹, Martín LOBOS¹,
Yamila MINETTI², Antonella BARBERO¹

¹ Departamento de Matemática/FCEyN/UNLPam, ²FCH/UNLPam
astudillo@exactas.unlpam.edu.ar, silviabast@exactas.unlpam.edu.ar

RESUMEN

El grupo de investigación GrIDIE se enfoca, desde 2018, en la investigación y el desarrollo de propuestas innovadoras que impacten en los procesos de enseñanza y de aprendizajes de informática. Particularmente, se hace foco en la inclusión de la programación en los niveles medio y superior de la educación formal.

En este artículo se presentan avances realizados en cada una de las líneas de acción del proyecto de investigación “Incorporación de Estrategias innovadoras en los Procesos de Enseñanza y de Aprendizajes de Informática”.

Palabras clave: Didáctica de la programación, calidad en EDI, programación tangible, Uso educativo de Telegram, Estilos de aprendizaje

CONTEXTO

El grupo de investigación GrIDIE¹ (Grupo de Investigación y Desarrollo en Innovación Educativa) se enfoca, desde 2018, en investigar la “Incorporación de estrategias innovadoras en los procesos de enseñanza y de aprendizajes de informática”. Este proyecto del GrIDIE, tiene seis líneas de investigación: (i) la medición y evaluación de Entornos de Desarrollo Integrado (EDI) para robótica educativa, (ii) la definición de criterios de evaluación que permitan identificar las posibilidades de la programación tangible para el aprendizaje de nociones básicas de programación, (iii) la revisión/evaluación de propuestas didácticas para la enseñanza/aprendizaje de la programación en función de su enfoque pedagógico/didáctico, (iv) la implementación de una propuesta

didáctica para el aprendizaje de la programación en el ámbito universitario, (v) el uso de Telegram como herramienta didáctica dentro de la asignatura Introducción a la Computación y (vi) la identificación de estilos de aprendizaje de los estudiantes que cursan Introducción a la Computación.

El proyecto es financiado por la Facultad de Ciencias Exactas y Naturales (FCEyN - UNLPam) y fue aprobado para el período 2018-2021 por RCD 27/18 y extendido hasta 2022 por RCD 514/21 FCEyN.

1. INTRODUCCIÓN

Actualmente hay una ingente cantidad de propuestas que se enfocan en la inclusión de la programación en la educación formal. Esto, según Popat & Starkey (2019) se puede atribuir a tres razones: (i) preparar a la fuerza laboral del futuro, (ii) estudiantes que aprenden a ser "productores" de innovación y (iii) situar al aprendizaje de la programación como competencias del siglo XXI. También, la didáctica de la programación comenzó a permear fuertemente en las aulas universitarias.

En Argentina se han impulsado distintas normas que apoyan el proceso: Ley N. 26.206, Ley de Educación Nacional en 2006, el Decreto 459/2010 en 2010, la Resolución CFE N. 263 de 2015 y la Resolución N. 1536-E del MED en 2017 y en 2018, la aprobación de los Núcleos de Aprendizaje Prioritarios para educación digital, programación y robótica.

A lo anterior se suman iniciativas a nivel nacional, con el propósito de que estudiantes de distintos niveles educativos aprendan a

¹ Facultad de Cs Exactas y Naturales, UNLPam

programar (Astudillo et al., 2019; Astudillo & Bast, 2020).

La motivación para avanzar con la investigación en la temática radica en que, más allá de los importantes avances que se han dado en los últimos tiempos a nivel nacional, la enseñanza y el aprendizaje de las Ciencias de la Computación (CC) continúa siendo tema de investigación y desarrollo a nivel nacional e internacional.

Introducción a la Computación (IC), y el Taller de Introducción a la Programación (TIP) (para ingresantes), son ámbitos donde se implementan la mayoría de las propuestas que surgen de las líneas de investigación.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

La hipótesis de trabajo del proyecto se basa en que es posible definir estrategias innovadoras para la enseñanza y el aprendizaje de temas de la programación, utilizando TIC, que impacten positivamente en la motivación de los estudiantes y, en los diseños didácticos de los docentes. Para ello se ha avanzado en seis líneas de trabajo:

La medición y evaluación de EDI

La motivación para avanzar en esta temática radica en que el TIP incluye el uso de EDI para fortalecer el concepto de secuencia, alternativa condicional, repetición y nociones de variables.

Esta línea de trabajo busca identificar, categorizar y evaluar distintos EDI para la enseñanza de robótica educativa, partiendo de la pregunta ¿Qué EDI de robótica educativa cuenta con la mayor flexibilidad para adaptarse a la propuesta didáctica del TIP?,

Dado que los EDI son desarrollos de software, se usa una metodología de evaluación de calidad, aplicando criterios establecidos por estándares internacionales, tal como la norma ISO 25010:2011. Para llevar a cabo el proceso de evaluación se hace uso de la estrategia denominada *Goal-Oriented Context-Aware Measurement and Evaluation* (Olsina et al, 2008).

Evaluación de posibilidades de la programación tangible

Existen lenguajes de programación textuales que pueden derivar muchas veces en errores de sintaxis y requieren un esfuerzo extra por quien debe aprender a utilizarlo. También existen en menor medida, lenguajes de programación del tipo icónico. En cualquier caso es necesario poseer conocimientos mínimos. Aquí es donde la incorporación de las interfaces de usuario tangibles refleja diversos beneficios. Esta se realiza a través de un teclado, mouse, pantalla y bloques tangibles de programación que representan algunas instrucciones en particular (Suzuki & Kato, 1993).

En esta investigación se busca responder: ¿Cuáles son las bases teóricas y/o enfoques que sustentan la Programación Tangible (PT)?, ¿Qué experiencias existen que utilizan PT en contextos educativos en los últimos diez años? ¿Cuáles son las características de la utilización de PT en contextos educativos? ¿Qué conjunto de criterios permitirían un análisis de las experiencias con PT?

Propuestas didácticas para la enseñanza programación

Conocer el sustento pedagógico de las propuestas didácticas es central para entender cuál es su alcance educativo. En esta línea se parte de las preguntas ¿Cuáles son las propuestas a nivel nacional que se enfocan en la enseñanza y el aprendizaje de la programación? ¿Cuál es el modelo pedagógico-didáctico que las sustenta? ¿Cómo conciben la enseñanza de la programación? ¿Qué aprendizajes se fomentan en los estudiantes con su implementación?

Durante 2018 y 2019, se identificaron y categorizaron distintas propuestas de enseñanza de la programación (Astudillo & Bast, 2020, Astudillo et. al, 2019). En esta segunda revisión el objetivo es profundizar la categorización de las propuestas didácticas en función de su enfoque pedagógico-didáctico, seleccionando las experiencias de educación superior y enfocando el análisis en los componentes didácticos presentados, el posicionamiento docente ante la enseñanza de

la programación, la elaboración de estrategias didácticas y materiales educativos, el rol asignado al estudiantado, la evaluación de los aprendizajes.

Resolver problemas y programar

Desde la cátedra IC, se han implementado y evaluando distintas estrategias para mejorar el acercamiento de los estudiantes a las nociones básicas de programación. Las propuestas se centraron, mayoritariamente, en los conceptos básicos de la programación, dejando en un segundo plano la resolución de problemas. Es por esto que surge la pregunta ¿Cómo enseñar una estrategia de resolución de problemas, transversal a la cursada, que le permita a los estudiantes apropiarse de la misma y que funcione como una herramienta que les permita abordar la solución de las actividades propuestas desde la asignatura?

Se plantea, entonces, la incorporación de estrategias de resolución de problemas desde el inicio de la cursada. Esta propuesta didáctica, basada en parte en la propuesta de la Fundación Sadosky (Martínez López, 2016) y en la resolución de problemas de Thomson (Thompson, 1997), se enfoca en la abstracción, en la división en subproblemas, el uso de procedimientos y la legibilidad.

Cada problema de la guía práctica se inicia con la pregunta ¿Qué hay que hacer? Se espera aquí que puedan definir en una oración corta cuál es el problema a resolver (Comprender el problema). Con base en las respuestas se asocian los procedimientos o partes del problema a nodos de un mapa mental (Diseñar el programa). Esto funciona como una plan para ir a programar (Escribir el programa). La estrategia de solución se refleja en el código fuente del programa a través de comentarios y luego se comienza a codificar. Finalmente, y con la solución inicial (del problema o parte de él) se revisa la solución en busca de mejoras (Mirar hacia atrás) (Astudillo, Bast & Minetti, 2021).

Uso de Telegram como herramienta didáctica

La retroalimentación es un aspecto importante tanto en el proceso de aprendizaje. Apunta desde el docente a generar un andamiaje para que sus estudiantes logren sus objetivos y mejoren sus procesos de aprendizaje. Además, permite que los estudiantes analicen sus acciones, ofreciendo información sobre los logros, los desafíos y cómo mejorar (Anijovich, 2019).

En el caso de IC, en pre-pandemia, la comunicación con los estudiantes se daba en los encuentros presenciales y los foros. A partir de 2020, ya en pandemia, fue imposible hacer el seguimiento y retroalimentación presencial de las producciones. Se avanzó entonces en una comunicación alternativa, así surge el grupo de Telegram de la asignatura. A partir del intenso intercambio que se produjo, desde el grupo de investigación, surge la pregunta ¿Cuáles son las posibilidades de uso de la red social Telegram para andamiar las actividades de la cátedra?

Estilos de aprendizaje

Indagar en los Estilos de Aprendizaje (EA) de los estudiantes tiene relevancia para el diseño de la enseñanza y en la personalización de los aprendizajes. Felder & Silverman (1988) afirman que el aprendizaje de un estudiante se rige, por una parte, por habilidades propias que han sido adquiridas previamente, pero también por su estilo de aprendizaje y la compatibilidad de éste con el estilo de enseñanza del profesor.

Existen muchos enfoques teóricos sobre EA, en esta investigación se tomará el de Honey & Mumford (1986) que definen cuatro categorías de estudiantes: activos (necesitan de la experiencia directa), reflexivos (basan su aprendizaje en la observación y recogida de datos), teóricos (necesitan de la conceptualización abstracta y formación de conclusiones) y pragmáticos (basado en la búsqueda de aplicaciones prácticas).

Cabe preguntarse, entonces: ¿Cuáles son los estilos de aprendizaje que pueden identificarse en los estudiantes que cursan IC?

3. RESULTADOS

A continuación se detallan los avances alcanzados en cada una de las distintas líneas de trabajo.

La medición y evaluación de EDI

Se ha desarrollado el modelo de calidad y se generó la totalidad de las métricas que se usarán para futuras mediciones y evaluaciones. Queda como trabajo futuro aplicar las mediciones, diseñar los indicadores y aplicarlos para obtener el resultado de la evaluación de calidad.

Definición de criterios de evaluación para identificar las posibilidades de la PT

Se inicia esta investigación con una revisión sistemática de literatura, y se utiliza la metodología de Kitchenham et al. (2009).

Se definieron las preguntas de investigación, el criterio de búsqueda y los criterios de inclusión y exclusión. Se realizó la búsqueda sobre IEEE Xplore y ACM Digital Library (Castro et al., 2021). Actualmente se extendió la búsqueda a ScienceDirect y SCOPUS y se está llevando adelante el proceso de inclusión y exclusión, y análisis de los artículos.

Revisión/evaluación de propuestas didácticas para la enseñanza/aprendizaje de la programación

Se analizaron 19 experiencias que abordan estrategias para la enseñanza de la programación en Nivel Universitario. En función de este análisis se puede identificar que, en general, el propósito está enfocado en mejorar algunas estrategias didácticas y el objetivo en desarrollar en los estudiantes, principalmente, el pensamiento computacional y los conceptos básicos de programación. La propuesta de enseñanza intenta superar el modelo tradicional, a partir de la utilización del aprendizaje basado en problemas y por descubrimiento/indagación, y planteando una interacción fluida y constante entre docentes y estudiantado.

Se busca dar seguimiento de las 19 experiencias analizadas y extender la

búsqueda, con la finalidad de ampliar y profundizar el abordaje de estrategias para la enseñanza de la programación.

Resolver problemas y programar

Durante 2020 y 2021 se desarrolló e implementó la propuesta didáctica en IC. Ésta debió compatibilizarse con el uso del lenguaje Pascal y el EDI Lazarus. Del análisis de resultados surge que la mayoría de los estudiantes se apropian de la metodología y que la propuesta tiene potencial para andamiar el proceso de resolución de problemas.

Como trabajos futuros, se rediseñará el material de cátedra para adaptarlo a la propuesta y se profundizará en el análisis de las producciones para observar la evolución en la apropiación de la propuesta en 2022.

Análisis del uso de Telegram, dentro de la asignatura Introducción a la Computación

Inicialmente se identificó el tipo de información que se mueve a través de la red, se realizó un primer análisis de los datos, se generó una aplicación para el registro y se completó el relevamiento de las interacciones.

Actualmente se han relevado 1108 interacciones. Donde por ejemplo: 21 son avisos de la cátedra y 658 son discusiones acerca de dudas sobre las Guías Prácticas.

Como trabajo futuro queda seguir avanzando con el relevamiento en 2022, identificar niveles de interacción, tendencias en conceptos/ejercicios consultados, hábitos de estudio, con el objetivo de mejorar y retroalimentar la práctica docente.

Identificación de estilos de aprendizaje de los estudiantes del que cursan IC

En esta línea se analizaron distintas teorías sobre EA y sus cuestionarios de evaluación. Para identificar los EA, se seleccionó el cuestionario de Honey Alonso CHAEA (Alonso, Gallego & Honey, 1997). Éste consta de 80 afirmaciones dividido en cuatro secciones de 20 ítems correspondientes a los cuatro estilos de aprendizaje (activo, reflexivo, teórico y pragmático). Además cuenta con un baremo, para evaluar los resultados, enfocado en CC.

La recolección de datos se llevará a cabo en marzo de 2022. En paralelo con el análisis de los datos se llevará adelante la definición del estado de un arte sobre EA en CC a nivel de Iberoamérica.

4. FORMACIÓN DE RECURSOS HUMANOS

En este proyecto participan un investigador formado y tres investigadores en formación. Dos de ellos avanzando sobre sus tesis de doctorado y un tercero en su tesis para alcanzar el grado de magíster.

El proyecto cuenta con tres graduados que se inician en la investigación. Uno de ellos desarrollando su trabajo en el marco de una beca de estímulo a las vocaciones científicas (concluída en 2021) y otro avanzando en su trabajo final de especialización con una beca de iniciación a la investigación.

Asimismo, se incorporaron al proyecto cuatro estudiantes de grado como asistentes de investigación. Una enfocada en el procesamiento de datos de Telegram y las otras tres en la revisión del estado del arte de estilos de aprendizaje en estudiantes de CC en el contexto iberoamericano. Esto lo hacen en el marco de las actividades para alcanzar el grado de Profesoras en Computación.

5. BIBLIOGRAFÍA

Alonso, C. M., Gallego, D. J., & Honey, P. (1997). *Los estilos de aprendizaje: Procedimientos de diagnóstico y mejora*. Mensajero Bilbao, España.

Anijovich, R., Mora, S., & Luchetti, E. (2009). *Estrategias de enseñanza: Otra mirada al quehacer en el aula* (Vol. 1). Aique.

Astudillo, G. J., Bast, S. G., & Minetti, Y. (2021). *Hacia una propuesta didáctica para la enseñanza de la programación*. TE&ET 2021.

Astudillo, G. J., & Bast, S. (2020). *Enseñanza y aprendizaje de programación. Hacia un estado del arte*. VEC, 11(20), 138-155.

Astudillo, G. J., Bast, S. G., Segovia, D., & Castro, L. (2019). *Revisión de propuestas para la enseñanza de la programación*. TE&ET 2019.

Castro, L. J., Artola, V., Astudillo, G. J., & Bast, S. G. (2021). *Hacia un estado del arte sobre programación tangible*. Actas JADICC, 45-53.

Felder, R. M., & Silverman, L. K. (1988). *Learning and teaching styles in engineering education*. Engineering education, 78(7), 674-681.

Honey, P., & Mumford, A. (1986). *The manual of learning styles* Peter Honey. Maidenhead.

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). *Systematic literature reviews in software engineering – A systematic literature review*. I&ST, 51(1), 7-15.

Martínez López, P. E. (2016). *Sugerencias para el dictado del curso La programación y su didáctica. Método Program*. AR. Fundación Sadosky.

Popat, S., & Starkey, L. (2019). *Learning to code or coding to learn? A systematic review*. Computers & Education, 128, 365-376.

Suzuki, H., & Kato, H. (1993). *Algoblock: A tangible programming language, a tool for collaborative learning*.

Thompson, S. (1997). *Where do I begin? A problem solving approach in teaching functional programming*. En H.Glaser, P.Hartel, & H.Kuchen (Eds.), *Programming Languages: Implementation, Logics, and Programs* (pp. 323-334). Springer.

Olsina, L., Papa, F., & Molina, H. (2008). *How to measure and evaluate web applications in a consistent way*. In *Web Engineering: Modelling and Implementing Web Applications* (385-420). Springer, London.

Verificación Formal de Software en Sistemas de Big DATA

Fernando Asteasuain –Rafael Aragón – Luciana Rodriguez Caldeira- Nicolás Granata - Nahuel Patera-
Pablo Gamboa – Hang Shao Feng
Contacto Principal: fernando.asteasuain@uai.edu.ar
Centro de Altos Estudio CAETI – Universidad Abierta Interamericana
Emails: <nombre.apellido>@uai.edu.ar

Resumen

La Ingeniería de Software debe evolucionar para poder enfrentar los desafíos de un mundo moderno hiper conectado y con grandes volúmenes de información y datos disponibles para ser analizados. En este contexto, nuevas disciplinas como la denominada Ciencia de Datos [6] han surgido recientemente. Para llevar adelante estos desafíos se deben contar con herramientas para la verificación formal de sistemas basados en BIG DATA que cuentan con una fuerte interacción con áreas de la Inteligencia Artificial como Aprendizaje Automático [1, 2, 3, 4] para poder mantener los estándares esperados de rigurosidad y calidad.

Esta evolución requiere de novedosas técnicas para componer un sistema a través de sus múltiples aristas, junto con herramientas que sean eficaces pero también lo suficientemente flexibles y expresivas [5].

La presente investigación pretende dar un paso en pos de enfrentar este desafío, combinando técnicas de verificación formal con la Inteligencia Artificial, en especial con la teoría de juegos, la síntesis de comportamiento y el aprendizaje automático.

Palabras claves: *Big Data, Verificación Formal, Síntesis de Comportamiento.*

Contexto

La presente investigación se encuentra enmarcado dentro del proyecto “Framework para el Desarrollo de Software mediante Modularización Avanzada” desarrollado en el

Centro de altos Estudios CAETI de la Universidad Abierta Interamericana.

El proyecto se va renovando anualmente, siendo este el sexto año consecutivo del desarrollo de la investigación.

El grupo de investigación está constituido por un investigador formado, dos en formación, un estudiante de maestría, dos estudiantes de grado haciendo sus tesis de licenciatura y un estudiante de grado haciendo la práctica profesional supervisada dentro del marco del proyecto. El proyecto es financiado 100% por la Universidad Abierta Interamericana.

La composición del grupo de investigación ha sido estable en estos años, lo cual es fundamental para poder profundizar en las líneas de investigación. Si existe una dinámica más fluida en los estudiantes de grado, ya que la mayoría al terminar sus tesis de grado o práctica profesional opta por trabajos en la industria. Sin embargo, la posibilidad de continuar con tesis de Maestría/Doctorado es una opción que se ha logrado capitalizar en algunas ocasiones. En particular, para el presente año contamos con un estudiante de posgrado.

Existe también la posibilidad de transferencia y servicios al sector industrial y productivo. Es importante mencionar que las líneas de Investigación del proyecto tienen impacto en áreas prioritarias del *Plan Nacional de Ciencia y Técnica 2020* como AgroIndustria, Biodiversidad e Innovación Productiva.

1. Introducción

La creciente demanda de sistemas basados en datos implica un gran esfuerzo para la comunidad de la Ingeniería de Software para poder validar y verificar el comportamiento esperado de sistemas en este dominio. Las características de los sistemas de BIG DATA se pueden resumir en lo que se conoce como las cinco “V”: Variedad, Velocidad, Volumen, Valor y Veracidad [17, 18]:

- 1) Volumen: la cantidad creciente de información obtenida.
- 2) Variedad: Diversas y heterogéneas fuentes y tipos de dato.
- 3) Velocidad: para la manipulación y adquisición de datos, streaming en tiempo real, y datos de tiempo variable.
- 4) Valor: El peso de cada dato dentro de toda la información disponible.
- 5) Veracidad: confianza en los datos obtenidos y su procesamiento

Sin embargo, en la aplicación de técnicas y herramientas tradicionales de la Ingeniería de Software en este tipo de sistemas se han encontrado algunas debilidades como la falta de modelado y diseño y dificultades a la hora de validar y verificar el comportamiento esperado [9,10,11,12,13]. Incluso algunos autores han denominado al big data como “no testeable”, proponiendo técnicas como testing metamórfico [20,21,22]. La verificación formal es el área que más necesita crecer dentro de la Ingeniería de Software [13]. Un enfoque en este sentido son los estándares de calidad establecidos para cuantificar la calidad de los dato [19]. Sin embargo, según lo reportado en [13] solo 2 de los casi 200 trabajos analizados vinculando Ingeniería de Software y BIG DATA se enfocan en temas de verificación formal. La mayoría de estos enfoques se concentran en sólo dos de las cinco “V”: Velocidad y Volumen, buscando mejorar la performance y el tamaño de exploración de los sistemas a analizar. Ejemplos exitosos han sido versiones paralelas o distribuidas de herramientas como model checkers. Sin embargo, las otras tres

“V”, Variedad, Valor y Veracidad no han recibido la atención necesaria. Dichas características necesitan formalismos y notaciones que sean expresivos y flexibles, para poder lidiar con un volumen gigantesco y poco estructurado información de información y datos.

Un enfoque novedoso en la búsqueda de formalismos más poderosos puede darse desde el área de Síntesis de comportamiento [7,8]. Se llama síntesis de comportamiento al proceso de obtener automáticamente un controlador de un sistema a partir de su especificación de manera que por construcción se garantice que las propiedades que describen el comportamiento del sistema son satisfechas.

Un controlador es en esencia un autómatas que recibe información de sensores, la procesa, y envía instrucciones a actuadores. La construcción de un controlador se realiza a través de técnicas de la Inteligencia Artificial y la teoría de juegos: se trata de obtener una estrategia ganadora para nuestro sistema que no importe que acciones se disparen desde el ambiente, se podrán cumplir los objetivos de comportamiento propuestos. Este esquema es clásico en entornos de sistemas abiertos como Robótica o Internet de las cosas.

La inmensidad de este desafío requiere entonces la necesidad combinar técnicas como la síntesis de comportamiento y el aprendizaje automático con herramientas tradicionales de verificación formal como model checking [14,15,16].

Una solución a los mencionados problemas tendrá impacto en las mencionadas áreas consolidando métodos, técnicas y herramientas de la Ingeniería de Software atacando cada una de las cinco “V” que caracterizan a Big Data.

2. Líneas de Investigación y Desarrollo

En el presente trabajo se explorarán las siguientes líneas de investigación:

- Aplicación de métodos formales de Ingeniería de Software al proceso de aprendizaje automático.
- Aplicación de métodos formales de Ingeniería de Software en redes neuronales.
- Aplicación de métodos formales de Ingeniería de Software en sistemas de BIG DATA.
- Aplicación de métodos formales de Ingeniería de Software en sistemas basados en Internet de las Cosas.
- Explorar la Síntesis de Comportamiento como un mecanismo poderoso para especificar comportamiento en sistemas orientados a datos.
- Analizar posibilidad de mejoras en los algoritmos para obtener controladores y en la expresividad de los lenguajes de especificación.
- Continuar y profundizar el desarrollo de herramientas de software que den aplicabilidad a los conocimientos adquiridos.
- Reforzar la interacción con model checkers distribuidos y/o paralelos. Las soluciones de software distribuidas y/o paralelas son claves para la optimización en tiempo que requieren los sistemas de BIG DATA.
- Contribuir en facilitar el proceso de verificar sistemas de BIG DATA.
- Potenciar y fusionar la cada vez más fuerte interacción entre la Ingeniería de Software y la Inteligencia Artificial.
- Expandir las nociones de estándares de calidad sobre manejo y manipulación de datos.
- Obtener demostraciones formales de la correctitud y completitud de los procesos involucrados.
- Expandir las nociones de testing metamórfico a la especificación de comportamiento. Esto implica buscar

propiedades “metamórficas” para la verificación de sistemas en este dominio.

3. Resultados Obtenidos/Esperados

El objetivo principal del presente proyecto de Investigación es aplicar nuevos métodos y técnicas rigurosas y formales de la Ingeniería de Software al procesamiento de grandes volúmenes de datos conocido como sistemas Big Data.

El desafío involucra la interacción de diversas áreas como Inteligencia Artificial, Redes Neuronales, Aprendizaje Automático, Síntesis de Comportamiento o el Procesamiento Dinámico de Información. Se combinarán técnicas de aprendizaje automático y síntesis de comportamiento para aplicar todo el potencial de la Ingeniería de Software a sistemas orientados a BIG DATA y ciencia de datos.

Como objetivos específicos podemos mencionar:

- a) Analizar los desafíos que requieren las técnicas, procesos y herramientas de Ingeniería de Software para lidiar con los sistemas de Big DATA.
- b) Razonar sobre la manera de poder validar el comportamiento esperado en sistemas de BIG DATA.
- c) Explorar sistemas paralelos de verificación formal para atacar los problemas de performance en sistemas con grandes volúmenes de información.
- d) Modelar, abstraer y razonar para la descripción de eventos de alto nivel de interés en áreas de Inteligencia Artificial.
- e) Aplicar herramientas y técnicas formales de la Ingeniería de Software para el desarrollo de sistemas de Big Data.
- f) Divulgar los resultados de la investigación en congresos y revistas científicas de interés para los temas de la investigación.

g) Consolidar los recursos humanos en inicios de las tareas de investigación como estudiantes avanzados de la carrera, mediante la participación activa en el proyecto o mediante la realización de tesis de final de carrera de grado.

Los resultados esperados incluyen:

- Consolidar y profundizar técnicas formales de la Ingeniería de Software en un área de aplicación de vanguardia como Big Data e Internet de las Cosas y Ciencia de Datos.
- Dirección de tesis de licenciatura, de maestría y supervisión de prácticas profesionales (PPS).

4. Formación de Recursos Humanos

El presente proyecto buscará potenciar a los investigadores en formación y de posgrado. En cuanto a los estudiantes de grado, se buscará en primer lugar que finalicen su tesis de grado o su práctica profesional. Sin embargo, se los pretende motivar como para que puedan continuar dentro del proyecto mediante la realización de algún posgrado.

En este sentido, una táctica que ha resultado exitosa fue la presencia activa de los estudiantes de grado en los congresos científicos del área, ya sean virtuales o presenciales.

5. Bibliografía

1. ALPAYDIN, Ethem. Machine learning. MIT Press, 2021.
2. WANG, Ping; LI, Yan; REDDY, Chandan K. Machine learning for survival analysis: A survey. *ACM Computing Surveys (CSUR)*, 2019, vol. 51, no 6, p. 1-36.
3. HAN, Rui; JOHN, Lizy Kurian; ZHAN, Jianfeng. Benchmarking big data systems: A review. *IEEE Transactions on Services Computing*, 2017, vol. 11, no 3, p. 580-597.
4. ELSHAWI, Radwa, et al. Big data systems meet machine learning challenges: towards big data science as a service. *Big data research*, 2018, vol. 14, p. 1-11.
5. VARDI, Moshe Y. Branching vs. linear time: Final showdown. *International conference on tools and algorithms for the construction and analysis of systems*. Springer, Berlin, Heidelberg, 2001. p. 1-22.
6. PROVOST, Foster; FAWCETT, Tom. Data science and its relationship to big data and data-driven decision making. *Big data*, 2013, vol. 1, no 1, p. 51-59
7. S. Maoz and Y. Sa'ar. Aspectl: an aspect language for ltl specifications. In *AOSD*, pages 19-30. ACM, 2011.
8. N. Dippolito, V. Braberman, N. Piterman, and S. Uchitel. Synthesising nonanomalous event-based controllers for liveness goals. *ACM Tran*, 22(9), 2013.
9. Embley, David W., and Stephen W. Liddle. "Big data—conceptual modeling to the rescue." *International Conference on Conceptual Modeling*. Springer, Berlin, Heidelberg, 2013.
10. Arndt, Timothy. "Big Data and software engineering: prospects for mutual enrichment." *Iran Journal of Computer Science* 1.1 (2018): 3-10.
11. Camilli, Matteo. "Coping with the State Explosion Problem in Formal Methods: Advanced Abstraction Techniques and Big Data Approaches." (2015).
12. Hummel, Oliver, et al. "A collection of software engineering challenges for big data system development." *2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. IEEE, 2018.
13. Kumar, Vijay Dipti, and Paulo Alencar. "Software engineering for big data projects:

Domains, methodologies and gaps." 2016 IEEE International Conference on Big Data (Big Data). IEEE, 2016.

14. SOURI, Alireza, et al. Formal verification of a hybrid machine learning-based fault prediction model in Internet of Things applications. *IEEE Access*, 2020, vol. 8, p. 23863-23874.

15. URBAN, Caterina; MINÉ, Antoine. A review of formal methods applied to machine learning. *arXiv preprint arXiv:2104.02466*, 2021.

16. W. Nam and H. Kil, "Formal Verification of Blockchain Smart Contracts via ATL Model Checking," in *IEEE Access*, vol. 10, pp. 8151-8162, 2022, doi: 10.1109/ACCESS.2022.3143145.

17. CAPPÀ, Francesco, et al. Big data for creating and capturing value in the digitalized environment: Unpacking the effects of volume, variety, and veracity on firm performance. *Journal of Product Innovation Management*, 2021, vol. 38, no 1, p. 49-67.

18. NAEEM, Muhammad, et al. Trends and future perspective challenges in big data. En *Advances in Intelligent Data Analysis and Applications*. Springer, Singapore, 2022. p. 309-325.

19. ISO 25012 (2008), "Ingeniería de software - Requisitos de calidad y evaluación de productos de software (SQuaRE) - Modelo de calidad de datos", Disponible en: <https://www.iso.org/obp/ui/es/#iso:std:iso-iec:25012:ed-1:v1:en>, consultado en febrero 2022.

20. CHEN, Tsong Yueh, et al. Metamorphic testing: A review of challenges and opportunities. *ACM Computing Surveys (CSUR)*, 2018, vol. 51, no 1, p. 1-27.

21. SEGURA, Sergio, et al. A survey on metamorphic testing. *IEEE Transactions on software engineering*, 2016, vol. 42, no 9, p. 805-824.

22. DING, Junhua; ZHANG, Dongmei; HU, Xin-Hua. A framework for ensuring the quality of a big data service. En *2016 IEEE International Conference on Services Computing (SCC)*. IEEE, 2016. p. 82-89.

Aproximación a un método de cálculo de viabilidad del despliegue de sistemas de software

Leandro Moreno¹, Marisa Panizzi^{1,2}, Rodolfo Bertone³

¹ Escuela de Posgrado. Universidad Tecnológica Nacional. Universidad Tecnológica Nacional. Facultad Regional Buenos Aires. Medrano 951 (C1179AAQ), C.A.B.A, Argentina.

² Departamento de Ingeniería en Sistemas de Información. Universidad Tecnológica Nacional. Facultad Regional Buenos Aires. Medrano 951 (C1179AAQ), C.A.B.A, Argentina.

³ Instituto de Investigación en Informática (III-LIDI). Facultad de Informática. Calles 50 y 120 - La Plata - Bs. As. – Argentina.

lm.moreno1991@gmail.com; marisapanizzi@outlook.com; pbertone@lidi.info.unlp.edu.ar

Resumen

La línea I+D+I presentada se enmarca en la búsqueda del fortalecimiento del proceso de despliegue de sistemas de software en Pequeñas y Medianas Empresas (PyMES) de Argentina. Este tipo de empresas requieren cada vez de procesos de software sistematizados y controlados con el propósito de lograr un mayor nivel de competitividad en la industria del software a nivel nacional como internacional. Este tipo de empresas requieren de herramientas y métodos que les permitan la mejora y estabilización de sus procesos. En esta línea de investigación y desarrollo para que las PyMES puedan anticipar los inconvenientes en su proceso de despliegue de sistemas de software, se propone el diseño de un método que les permita calcular la viabilidad de este. Antes de comenzar con el diseño del método para el cálculo de viabilidad del despliegue de sistemas de software, se realizó una revisión exploratoria de literatura existente de la cual se comparten los primeros hallazgos.

Palabras clave: procesos de software, proceso de despliegue, sistemas de software, viabilidad.

Contexto

La línea I+D+I que se reporta en este artículo es financiada parcialmente por un proyecto de investigación titulado “Estudio del proceso de implantación de sistemas informáticos en el contexto industrial de la República Argentina” (Código SIUTNBA0006576) de la Secretaría de Ciencia, Tecnología y Posgrado de la Universidad Tecnológica Nacional.

Cuenta con el asesoramiento científico del grupo de investigación de Ingeniería de Software del Instituto de Investigación en Informática (III-LIDI) de la UNLP.

Introducción

El despliegue de sistemas de software es uno de los procesos fundamentales dentro del ciclo de vida del desarrollo del software, de este proceso depende la aceptación del sistema de software por parte de los usuarios.

Al igual que el resto de los procesos involucrados, el despliegue de sistemas de software abarca actividades y prácticas que pueden fallar o poner en riesgo el éxito del proceso: la ausencia de componentes (externos), descargas incompletas y despliegues erróneos [1]. En consecuencia, a menudo los inconvenientes ocurridos en la fase de despliegue se transfieren y eventualmente resuelven como parte de la fase de

mantenimiento. En este sentido, las empresas pueden demorarse meses o años en lograr la estabilidad de un sistema finalizando el despliegue. Por este motivo, debe considerarse que un proceso eficiente de despliegue de software deberá ahorrar recursos financieros y humanos, es decir, en términos de costo y esfuerzo [2].

Asimismo, otro inconveniente que aporta complejidad e ineficiencia al momento de efectuar un despliegue de un sistema de software es la presencia de entornos heterogéneos y distribuidos y heterogéneos [3].

Por otra parte, este proceso no puede verse únicamente desde una óptica tecnológica, sino que se encuentra atravesado por la complejidad de la estructura organizacional existente, el cambio en la forma de trabajar de las personas y falta de experiencia y habilidades [4].

Teniendo en cuenta que la actividad de despliegue involucra las acciones de entrega, apoyo y retroalimentación [5], los mecanismos de validación y aseguramiento de la calidad deben involucrar necesariamente cada una de ellas, y contemplar el relevamiento, diseño, medición y validación al interior de cada fase.

A partir de la reflexión acerca de los artefactos y metodologías de evaluación de estos procesos de despliegue como insumo necesario para incorporar dentro de la metodología propia de una organización, surge esta línea de investigación que busca, en primera instancia, efectuar una exploración acerca del estado del arte en la materia.

Petersen *et al.* en [6] reconocen cuatro elementos centrales para la evaluación de proyectos de software, adicionales a la organización y al mercado. Estos elementos son:

1) Producto: el sistema de software desarrollado. Entre los aspectos que permiten entender y evaluarlo en contexto, se destacan:

la madurez, la calidad, el tamaño, el tipo de sistema, el nivel de personalización o adaptación y el lenguaje de programación.

2) Procesos: el flujo de trabajo del desarrollo. Dentro de los aspectos de análisis se encuentran las actividades involucradas, el flujo de trabajo y orden de ejecución de las actividades y los artefactos resultado de cada actividad o iteración.

3) Prácticas, herramientas y técnicas: en este grupo se comprenden los aspectos que brindan un abordaje sistematizado en la interacción con el objeto de estudio. Aquí destacan las herramientas CASE y las metodologías aplicadas.

4) Factor humano: este aspecto es también muy relevante, dado que tiene un gran impacto sobre el proceso de desarrollo. En este punto, debe atenderse a los roles asignados al proyecto de software y su experiencia.

5) La organización comprendida por la estructura, atendiendo al modelo jerárquico y flujo de comunicaciones, las certificaciones que obtiene a partir de la capacitación y la proyección (local, internacional).

6) El mercado: compuesto por los competidores y potenciales clientes, se ve afectado por el número de estos últimos, el segmento involucrado, la estrategia de ventas en el largo plazo y restricciones.

Estos aspectos analizados se considerarán para el diseño del método de cálculo de viabilidad de despliegue de sistemas de software.

Antes de comenzar con el diseño del método de cálculo de viabilidad del despliegue, se realiza una revisión exploratoria de la literatura. Los primeros hallazgos conllevan que es un tema de interés en el área y que, además es materia de estudio la búsqueda de la optimización para el éxito de los despliegues en un escenario de creciente

demanda de soluciones informáticas que requieren un ciclo continuo de adaptaciones.

El próximo paso del proceso de investigación consiste en la realización de un mapeo sistemático de la literatura (en inglés, o *Systematic Mapping Study* o SMS) [7] riguroso con el propósito de dar respuesta a las siguientes preguntas de investigación (PI):

PI1: ¿Qué instrumentos se utilizan en la evaluación de los procesos de despliegue de software?

PI2: ¿Qué dimensiones se consideran en las evaluaciones que se realizan sobre los procesos de despliegue de software?

PI3: ¿Qué tipos de investigaciones se encuentran en los artículos?

Actualmente se encuentra en desarrollo el protocolo de revisión de este SMS. Algunas de las consideraciones que han sido definidas por los investigadores, se detallan a continuación:

- Búsqueda automática en las librerías digitales IEEE Xplore, Scopus y ACM.
- Lectura de artículos de congresos y artículos de revistas.
- Período comprendido entre enero del año 2016 hasta junio del año 2021.
- Idioma inglés.

Además, se ha definido que para la clasificación de los tipos de investigación para los artículos se utilizará la propuesta por

Wieringa *et al.* [8].

Líneas de Investigación, Desarrollo e Innovación

La línea de I+D+I presentada consiste en el diseño de un método que permita calcular la viabilidad del proceso despliegue de sistemas de software. Una vez diseñado el método se

validará en la industria mediante un conjunto de estudios de caso [9].

Resultados y Objetivos

En esta línea de I+D+I en progreso respecto al fortalecimiento del proceso de despliegue de sistemas de software se han logrado una serie de resultados que se detallan a continuación:

- a) Académicos, se han logrado dos trabajos de especialidad en Ingeniería en Sistemas de Información, una tesis de Maestría en Ingeniería en Sistemas de Información defendida y otra en desarrollo y un trabajo de especialidad en Ingeniería en Sistemas de Información en desarrollo. Además, se ha depositado una tesis Doctoral en Ciencias Informáticas.
- b) Producción Científica: se ha presentado la línea de I+D+I en eventos científicos de alcance nacional (WICC¹ 2017, WICC 2019, WICC 2020, WICC 2021 CACIC² 2016, CACIC 2018, CACIC 2019, CACIC 2020, CACIC 2021 y en el ámbito internacional, CIACA³ 2017, SEKE⁴ 2017, CIBSE⁵ 2019, CIBSE 2020 CIBSE 2021 y ICAETT⁶ 2019. Además, se lograron tres publicaciones en Springer y en la revista *Brazilian Journal of Development*.
- c) Proyectos de Investigación: el PID UTNBA4347 titulado: "Impacto del factor peopleware en el proceso de implantación de sistemas informáticos" (período 2017-2019) que ha finalizado. El PID UTNBA6576 titulado: "Estudio del proceso de implantación de sistemas informáticos en el contexto industrial de la República Argentina" (período 2020-2022) que continúa en ejecución.
- d) Formación en investigación: el grupo de investigación se encuentra en un

¹ Workshop de Investigadores en Ciencias de la Computación

² Congreso Argentino de Ciencias de la Computación.

³ Conferencia Iberoamericana de Computación Aplicada.

⁴ International Conference on Software Engineering and Knowledge Engineering.

⁵ Congreso Iberoamericano en Ingeniería de Software.

⁶ Congreso Internacional sobre Avances en Nuevas Tendencias y Tecnologías.

proceso de aprendizaje constante de métodos de investigación de ingeniería de software experimental, revisiones sistemáticas [10], estudios de casos [11] y encuestas [12][13].

Formación de Recursos Humanos

El grupo se encuentra conformado por un director, dos tesis de Maestría, un tesista doctoral, un becario de investigación y un asesor científico-tecnológico.

Se estima la formación de un Doctor en Ciencias Informáticas y un Magister en Ingeniería en Sistemas de Información. En esta línea de I+D+I se incorpora un tesista de Maestría que actualmente se encuentra desarrollando como primera instancia su trabajo de especialidad.

Referencias

- [1] Jansen S., Brinkkemper S, (2006). Definition and validation of the key process of release, delivery and deployment for product software vendors: Turning the ugly duckling into a swan IEEE International Conference on Software Maintenance, ICSM, art. no. 4021334, pp. 166-175.
- [2] Subramanian N, (2017). The software deployment process and automation. *CrossTalk*, 30 (2), pp. 28- 34.
- [3] Tyndall J. (2012), Building an effective software deployment process. In Proceedings of the 40th annual ACM SIGUCCS conference on User services, pp. 109-114.
- [4] Reascos I., Carvalho J., Bossano S. (2019), Implanting IT Applications in Government Institutions: A Process Model Emerging from a Case Study in a Medium-Sized Municipality. In Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance, pp. 80-85.
- [5] Roger S. Pressman. (2010), Ingeniería del Software: un enfoque práctico. McGraw-Hill Interamericana de España.
- [6] Petersen, K. Wohlin, C. (2009), Context in industrial software engineering research. Third International Symposium on Empirical Software Engineering and Measurement.
- [7] Kitchenham, B., Charters, S. (2007), Guidelines for performing systematic literature reviews in software engineering, Keele University, EBSE-2007-01.
- [8] Wieringa, R., Maiden, N., Mead, N., Rolland, C. (2006), Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. *Requirements Engineering*, 11(1), pp 102-107}
- [9] Kitchenham B., Budgen, D., Brereton P. (2010), Evidence-Based Software Engineering and Systematic Reviews.
- [10] Runeson P., Höst M., Rainer A., Regnell B. (2012). Case study research in software engineering: guidelines and examples. Wiley Publishing, Hoboken.
- [11] Molléri J., Petersen K., Mendes E. (2020). An empirically evaluated checklist for surveys in software engineering. *Information and Software Technology*, 119.
- [12] Genero, M., Piattini, M., & Cruz Lemus, J. A. (2014), Métodos de investigación en Ingeniería del Software. Madrid: Ra-Ma Editorial y Publicaciones.

EXPERIENCIA del CIUDADANO: MEJORANDO la INTERACCIÓN en CONTEXTOS DIGITALES

Adriana MARTIN, Gabriela GAETAN, Viviana SALDAÑO, Claudia CARDOZO, Silvia VILLAGRA, Hernán SOSA, Mariela COLOMBRES

Grupo de Investigación y Formación en Ingeniería de Software (GIFIS)

Instituto de Tecnología Aplicada (ITA)

Universidad Nacional de la Patagonia Austral, Unidad Académica Caleta Olivia (UNPA-UACO)

adrianaelba.martin@gmail.com; gabriela gaetan@yahoo.com.ar; vsaldanio@gmail.com;

claudia_yoryi@yahoo.com.ar; svillagra@uaco.unpa.edu.ar; hassio_09@hotmail.com;

mariela_colombres@hotmail.com

RESUMEN

El impacto social y económico provocado por la pandemia del coronavirus COVID19, ha evidenciado en los dos últimos años, que el uso de contextos digitales es la clave para la supervivencia de las organizaciones públicas, de gobierno y privadas. El uso exitoso de contextos digitales requiere de personas que cambien sustancialmente la manera en que piensan y actúan al desempeñar sus roles: los empleados en una organización deben cambiar la forma en que ejecutan su trabajo, los consumidores deben cambiar la manera en que utilizan los productos y los ciudadanos deben redefinir la manera en que acceden e interactúan con los servicios públicos y de gobierno. Estas aseveraciones, conllevan a encontrar y sostener en el tiempo, formas efectivas para canalizar experiencias satisfactorias, independientemente del rol desempeñado por las personas como usuarios de contextos digitales.

El campo de Diseño de Experiencia de Usuario (UXD¹) fue creado para satisfacer este mandato, a los efectos de diseñar productos, servicios y soluciones basados en la experiencia del usuario. La eXperiencia de Usuario (UX²), es un enfoque que se refiere básicamente al conjunto de factores y elementos relativos a la interacción del usuario con un entorno, cuyo resultado es la generación de una percepción (experiencia) positiva o negativa de ese entorno. La UX es satisfactoria cuando la interacción con ese entorno, soluciona problemas reales a la gente real.

Los integrantes de Grupo de Investigación y Formación en Ingeniería de Software (GIFIS), han estado enfocados en mejorar los contextos de interacción para contribuir con experiencias de usuario satisfactorias, y específicamente desde el Proyecto de Investigación (PI) correspondiente al período 2020-2022, han estado principalmente enfocados en mejorar la experiencia del usuario ciudadano.

Palabras clave: *Experiencia del Ciudadano | Contextos Digitales | Diseño de Experiencia de Usuario (UXD) | Usabilidad | Accesibilidad | Realidad Aumentada | Geolocalización | Sistemas Aumentativos y Alternativos de Comunicación (SAAC)*

CONTEXTO

En el campo del diseño e implementación de productos, cuyo objetivo es proporcionar contexto digitales de acceso e interacción a los usuarios, es frecuente escuchar el uso de términos, tales como “diseño centrado en las personas” o “tecnologías centradas en las personas” y más recientemente “diseño de experiencia de usuario” a los efectos de destacar que las personas constituyen el origen y el destino de esos productos. En la práctica, es importante incorporar y aplicar estos términos como lo que son, herramientas para entender las características, retos, y necesidades de los usuarios [1][2].

En particular, resulta desafiante y motivador trasladar y aplicar el diseño de experiencia de usuario a la interacción del ciudadano, a los efectos de: (a) contribuir en el desarrollo de plataformas y productos que den respuesta a sus necesidades y, (b) ofrecer soluciones en vez de crearles nuevos problemas. Considerar el UXD

¹ UXD siglas en inglés: User eXperience Design.

² UX siglas en inglés: User eXperience.

en el sector de las organizaciones públicas y de gobierno, puede tener un rol clave para ofrecer servicios a la medida de sus ciudadanos.

Existen numerosos referentes a quien atribuir la concientización de la relevancia del usuario y su experiencia. En 1998, Ben Shneiderman [3], indiscutido referente en el campo de la Interacción Humano-Computador (HCI)³ publica “Las Ocho Reglas de Oro del Diseño de Interfaces”, donde propone aumentar la productividad del usuario a través de una interfaz gráfica que incluya elementos para aumentar la sensación de competencia, dominio y control del usuario sobre el sistema. Con anterioridad, Don Norman [2] populariza el término UX (usado por primera vez por Rob King, 1977), expresando que hablar sólo de interfaz y de Usabilidad es demasiado limitado para cubrir todos los aspectos de la experiencia de las personas con el sistema.

GIFIS, es un grupo de investigación perteneciente al Instituto de Tecnología Aplicada (ITA), de la Unidad Académica Caleta Olivia (UACO)⁴, Universidad Nacional de la Patagonia Austral (UNPA)⁵. El Proyecto de Investigación (PI) N° 29/B256 presentado en este artículo, se ejecutó en el Período: 2020-2022, y se denomina: “Contextos Digitales para Asistencia de los Ciudadanos: Enfoques de Experiencia de Usuario”, y fue dirigido por la Dra. Adriana Martín y la Mg. Gabriela Gaetán. Desde 2012, y en el marco de varios proyectos de investigación consecutivos y relacionados, los integrantes de GIFIS han estado trabajando desde el diseño, evaluación y desarrollo de productos Web y móviles, para incluir propiedades de usabilidad y accesibilidad Web que mejoren la experiencia de los usuarios. Este trabajo ha incluido diferentes dominios y perfiles de usuarios de interés.

En este contexto, el PI N° 29/B256 se planteó y se ejecutó como una continuación del PI 29/B222 (2018-2020): “Diseño y Evaluación de Experiencia de Usuario para Multi-Dispositivos”, aplicando estos antecedentes para mejorar y enriquecer la UX en contextos de interacción ciudadana. El próximo paso, es

proyectar los avances alcanzados en dos nuevos proyectos: el PI N° 29/B285 (2022-2025): “Desarrollo y Evaluación de Productos Web y Móvil centrados en la Experiencia de Usuario”, y el PDTS-UNPA (2022-2024): “Revalorizando el Patrimonio Histórico-Cultural de Caleta Olivia: Una Aplicación Móvil para Circuitos Turísticos y Educativos”.

1. INTRODUCCIÓN

Los avances en la tecnología han permitido ofrecer alternativas de interacción a las personas y cada vez son más los dominios (sociales, educativos, empresariales, gubernamentales, etc.), que ofrecen sus servicios a través de contextos digitales. Esta realidad, ha propiciado en las personas nuevas expectativas por los servicios que se proporcionan bajo esta modalidad, y revela la necesidad de mejorar la experiencia de interacción del ciudadano. En particular, ofrecer contextos digitales a la medida de las necesidades de las personas es una manera de demostrar que las organizaciones públicas y de gobierno pueden ofrecer servicios de calidad, a la vez que abren los caminos para futuras interacciones con los ciudadanos.

En los últimos años, varios países han emprendido este camino, siendo España [4] uno de los países de la Unión Europea con el mejor progreso, cuando se trata de digitalización. Actualmente, la preocupación del gobierno Español no sólo se enfoca en proporcionar servicios a través de buenos contextos digitales de interacción para con sus ciudadanos, sino que muy especialmente, en fomentar el uso de estos servicios. Este objetivo no se puede alcanzar con sólo ofrecer servicios correctos y eficaces, sino que además, se debe ofrecer algo “experiencial”.

La experiencia que genera en las personas el uso de productos y servicios digitalizados, parece haberse convertido en un tema crucial al momento de desarrollar contextos digitales para el soporte de estos productos y servicios. Algunos autores, tal como [1], consideran que es ahora el momento de pasar de una “economía de servicios” a una “economía de experiencias”, proponiendo tres componentes fundamentales para el UXD [1]:

³ HCI siglas en inglés: Human-Computer Interaction.

⁴ <<http://www.uaco.unpa.edu.ar/uaco/>>

⁵ <<http://www.unpa.edu.ar/>>

1. Pensamiento Sistémico: que ayuda a entender la experiencia del usuario, permitiendo descubrir y abordar el problema de raíz e idear soluciones para el usuario de manera integral;

2. Jerarquía de las Necesidades: que permite tratar a la experiencia del usuario más allá de la apariencia, para analizar cómo la solución satisface las necesidades del usuario y;

3. Método Científico: que permite la experimentación temprana poniendo a prueba suposiciones, mitigando riesgos, y definiendo las necesidades reales de los usuarios y el ecosistema, a partir de construir hipótesis y obtener evidencia empírica que ayude a definir el problema que estamos tratando de resolver y la efectividad de la solución que tenemos en mente.

Considerar la UX al desarrollo de productos y servicios digitalizados, es una herramienta muy poderosa, debido a que su objetivo es aumentar la satisfacción del usuario. La importancia de la experiencia que vivencian las personas con los productos y servicios digitalizados, no ha pasado desapercibida. Especialmente, en los últimos años, que se pone de manifiesto una tendencia a aplicar el UXD en contextos digitales para propiciar la interacción de los ciudadanos con los organismos públicos y de gobierno [5][6][7][8][9].

GIFIS ha comenzado a trabajar en esta perspectiva del usuario ciudadano a partir del presente PI N° 29/B256 (2020-2022), denominado: “Contextos Digitales para Asistencia de los Ciudadanos: Enfoques de Experiencia de Usuario”, utilizando como bases la investigación realizada y experiencias capitalizadas de la ejecución de los PI-UNPA 29/B167 (2014-2016), 29/B194 (2016-2018), y 29/B222 (2018-2020). Con el objetivo de contribuir a la creación de experiencias enriquecidas, las siguientes contribuciones surgen de la ejecución del presente PI [10][11][12][13][14][15][16][17][18][19].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El Objetivo General de investigación y desarrollo del PI N° 29/B256 es: “Aplicar Enfoques de Diseño y Evaluación de Experiencia en Contextos Digitales para Asistir

al Ciudadano”. Para alcanzar este Objetivo General, se establecieron los siguientes Objetivos Específicos:

OE1. Proponer estrategias de solución para problemas de UX en aplicaciones móviles (Apps) de Gobierno.

OE2. Diseñar un Ecosistema de Contenido para el dominio Universitario.

OE3. Construir soluciones de software orientadas a Patrimonio Cultural y Turismo.

OE4. Desarrollar un Sistema Alternativo y Aumentativo de Comunicación (SAAC) para el ámbito de la Salud.

Estos Objetivos Específicos, definen en GIFIS líneas de investigación y desarrollo, dentro de las cuales cada uno de los integrantes trabajó durante 2020-2021, para alcanzar sus objetivos de investigación y/o formación particulares.

3. RESULTADOS OBTENIDOS/ ESPERADOS

A continuación, se presenta una breve descripción sólo del trabajo realizado durante 2021, dentro de los Objetivos Específicos (OE) planteados en la Sección 2. (ver en [12], el trabajo realizado dentro de estos OE durante 2020). Dentro de OE2, se terminó el rediseño y se implementó el sitio Web de la UNPA-UACO⁶. A tal efecto: (i) se revisaron y reestructuraron todos los contenidos del sitio aplicando el diseño de estrategias de contenido, y el concepto de ecosistemas de contenido y, (ii) se incluyeron durante el proceso propiedades de usabilidad y accesibilidad Web. Inmerso en OE3, se trabajó en incorporar los beneficios de la Realidad Aumentada (RA) a contextos digitales del dominio Patrimonio Cultural y Turismo. En [14][17], se revisó el estado-del-arte de las plataformas y herramientas de RA, aprendiendo el uso y realizando pruebas, para determinar capacidades y, ventajas y desventajas de las mismas. En [18], se realizó una revisión de los estudios existentes sobre UX con contexto digitales destinados al Turismo Cultural con Realidad Aumentada y, se analizaron y recopilamos un conjunto de guías para favorecer la UX en este dominio. Este trabajo continuará en los nuevos proyectos con el objetivo de desarrollar soporte al patrimonio

⁶ <<https://www.uaco.unpa.edu.ar/>>

cultural de la provincia de Santa Cruz (zona norte), Patagonia Argentina, aplicando técnicas de RA y Geolocalización. Alineado a OE4, se trabajó sobre el concepto de Sistemas Aumentativo y Alternativo de Comunicación (SAAC), para desarrollar una aplicación móvil en dispositivo *Tablet*, la cual proveerá soporte específico al proceso de comunicación e interacción, entre los agentes de salud y los pacientes con discapacidad comunicativa. El desarrollo incorpora un conjunto de pictogramas del Centro Aragonés para la Comunicación Aumentativa y Alternativa (ARASAAC)⁷, que son propiedad del Gobierno de Aragón, distribuidos bajo Licencia Creative Commons BY-NC-SA⁸. A tal fin, se ha obtenido previamente el permiso de uso respectivo. La aplicación está dirigida a las áreas de guardia e internación del Hospital Zonal Caleta Olivia, para asistir en el proceso de atención de los pacientes con Necesidades de Comunicación Complejas (NCC). En [15], se presentan los avances del proyecto, el cual se encuentra en la etapa final de desarrollo e implementación. Este trabajo de investigación y desarrollo, fue presentado por su autor (integrante alumno de grado de GIFIS) en Jornadas de Accesibilidad⁹ y Encuentro de Tecnología¹⁰ UNPA-UACO.

4. FORMACIÓN DE RECURSOS HUMANOS

En la actualidad, GIFIS tiene 7 integrantes: 5 docentes-investigadores UNPA, 1 docente-investigador invitado externo UTN-FRTDF y, 1 alumno de grado UNPA.

Durante 2021, los integrantes docentes-investigadores de GIFIS han participado como evaluadores de Becas de Iniciación a la Investigación para Alumnos de Grado y Postgrado UNPA. También, han sido revisores de artículos para eventos científicos-técnicos (tal como la revista ICT-UNPA) y de tesis de grado y postgrado UNPA. Asimismo, un

integrante docente-investigador estuvo a cargo del dictado de un curso de postgrado en la Maestría en Informática y Sistemas UNPA.

Los integrantes docentes/investigadores de GIFIS, están en permanente actualización, para dar soporte a la ejecución de los proyectos de investigación y a la formación de sus integrantes. A los efectos de poder evaluar en productos concretos las contribuciones al UXD en el desarrollo de contextos digitales, en 2020 comenzaron pruebas de integración de herramientas para dar soporte a la RA y la Geolocalización. Las primeras pruebas se realizaron utilizando Unity¹¹ y luego, se migró a Android Studio¹², utilizando los lenguajes Java y Kotlin¹³. Esta actividad continuará durante 2022, con el objetivo de formar nuevos recursos que contribuyan a la concreción de los objetivos de investigación y desarrollo plasmados en los dos nuevos proyectos UNPA de GIFIS, los cuales comienzan su ejecución en 2022: el PI N° 29/B285 (2022-2025): “*Desarrollo y Evaluación de Productos Web y Móvil centrados en la Experiencia de Usuario*”, y el PDTs-UNPA (2022-2024): “*Revalorizando el Patrimonio Histórico-Cultural de Caleta Olivia: Una Aplicación Móvil para Circuitos Turísticos y Educativos*”.

5. AGRADECIMIENTOS

A la UNPA por el soporte al Proyecto de Investigación (PI) N°: 29/B256, Período: 2020-2022, Denominado: “*Contextos Digitales para Asistencia de los Ciudadanos: Enfoques de Experiencia de Usuario*”.

6. REFERENCIAS

- [1] Colima, V. Diseño de Experiencia de Ciudadano para la Transformación Digital. Inter-American Development Bank (IDB), Editorial Studio Analogous, 2021. Zamzami, K. F. The Key Service Feature of M-Government Based on Interactive User Experiences. IEEE Access. Vol. 7, DOI: 10.1109/ACCESS.2019.2924136, 2019.
- [2] Norman, D., Nielsen, J. The Definition of User Experience (UX) <<https://www.nngroup.com/articles/definition-user-experience/>> Accedido 2020.

⁷ <<https://arasaac.org/>>

⁸ <<https://creativecommons.org/licenses/by-nc-sa/3.0/es/>>

⁹ “II Jornadas de Discapacidad y Tecnologías”

<<https://www.uaco.unpa.edu.ar/las-ii-jornadas-de-discapacidad-y-tecnologias-se-realizaran-el-29-y-30-de-noviembre>>

¹⁰ “La Noche Mágica del Labtem después de la Pandemia”

<<https://www.lavanguardia.com.ar/nota/40504-se-realizo-la-muestra-anual-la-noche-magica-del-labtem-de-la-unpa-uaco-en-el-sum-del-centro-cultural/>>

¹¹ <<https://unity.com/>>

¹² <<https://developer.android.com/>>

¹³ <<https://kotlinlang.org/docs/android-overview.html>>

- [3] Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., and Elmqvist, N. *Designing the User Interface: Strategies for Effective Human-Computer Interaction: Sixth Edition*, Pearson, 2016.
- [4] Gobierno de España: Agenda 2030. Plan España Digital 2025. <https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/230720-Espa%C3%B1aDigital_2025.pdf> Accedido 2021.
- [5] Martínez de Lahidalga, J. Mejorar la Experiencia Digital del Ciudadano para Transformar los Servicios Públicos. <<https://ibermaticadigital.com/mejorar-la-experiencia-digital-del-ciudadano-para-transformar-los-servicios-publicos/>> Accedido 2021.
- [6] Skarlatidou, A., Ponti, M., Sprinks, J., Nold, C., Haklay, M., Kanjo, E. User Experience of Digital Technologies in Citizen Science. *Journal of Science Communication*. Vol. 18(1), pp.1-8, DOI: 10.22323/2.18010501, 2019.
- [7] Adarsha, A., Reader, K., Erban, S. User Experience, IoMT, and Healthcare. *Transactions on Human-Computer Interaction*. Vol. 11(4), pp. 264-27, DOI: 10.17705/1thci.00125, 2019.
- [8] EY Services. Driving Citizen Centricity: An Ecosystem to Help GCC Governments Sustainably Deliver Better Citizen Experiences. <[https://www.ey.com/Publication/vwLUAssets/EY-what-is-driving-gcc-governments-to-deliver-better-citizen-experiences/\\$FILE/EY-what-is-driving-gcc-governments-to-deliver-better-citizen-experiences.pdf](https://www.ey.com/Publication/vwLUAssets/EY-what-is-driving-gcc-governments-to-deliver-better-citizen-experiences/$FILE/EY-what-is-driving-gcc-governments-to-deliver-better-citizen-experiences.pdf)>, 2017. Accedido 2020.
- [9] Center for Public Impact. A BCG Foundation. Research & Conversation - Articles: Citizen Experience. <<https://www.centreforpublicimpact.org/>>, 2015-2018. Accedido 2021.
- [10] Cardozo, C., Martín, A., Saldaño, V., Gaetán, Gabriela. Una Propuesta para Mejorar la Experiencia de los Adultos Mayores con las Redes Sociales. *Revista Tecnología, Ciencia y Educación*. Centro de Estudios Financieros (CEF) /Universidad a Distancia de Madrid (UDIMA). N° 16, pp. 113-142, ISSN 2444-250X e ISSN-e 2444-2887, <<https://tecnologia-ciencia-educacion.com/index.php/TCE/article/view/445>>, 2020.
- [11] Vidal, P., Martín, A. Experiencia de Usuario + Web Responsivo: Un Estudio desde la Perspectiva de un Enfoque Integrado. *ICT-UNPA*, Vol. 12(1), pp. 49-75, ISSN: 1852-4516, DOI:<<https://doi.org/10.22305/ict-unpa.v12.n1.703>>, 2020.
- [12] Martín, A., Gaetán, G., Saldaño, V., Cardozo, C., Carrizo, A., Sosa, H., Piñero, R., Villagra, S. La Experiencia de Usuario en el Contexto de los Multi-Dispositivos. *XXII Workshop de Investigadores en Cs. de la Computación (WICC)*; pp. 532-542. ISBN: 978-987-3714-82-5. <<http://sedici.unlp.edu.ar/handle/10915/104046>>, 2020.
- [13] Carrizo, A., Gaetán, G. Gobierno Móvil en Argentina: Análisis de Experiencia de Usuario basado en Comentarios de Tiendas de Aplicaciones. *Revista ICT-UNPA*, vol. 12(1), pp.18-48, ISSN: 1852-4516. DOI: <<https://doi.org/10.22305/ict-unpa.v12.n1.702>>, 2020.
- [14] Cardozo, C.; Martín, A.; Saldaño, V., Gaetán, G. Análisis de Herramientas de Realidad Aumentada para Crear Experiencias de Turismo. *VI Encuentro de Investigadores, Becarios y Tesistas de la Patagonia Austral*, pp. 710-714, ISBN: 978-987-3714-88-7, 1ra. Ed Digital 2021.
- [15] Sosa, H., Martín, A., Saldaño, V., Gaetán, G. Desarrollando una Aplicación SAAC Móvil para el Hospital Zonal de Caleta Olivia. *VI Encuentro de Investigadores, Becarios y Tesistas de la Patagonia Austral*, pp. 726-730, ISBN 978-987-3714-88-7, 1ra. Ed. Digital 2021.
- [16] Martín, A., Gaetán, G., Saldaño, V., Cardozo, C., Vega, M., Carrizo, A., Villagra, S., Sosa, H. Enriqueciendo la Experiencia de Usuario en Dominios para Asistencia de los Ciudadanos. *VI Encuentro de Investigadores, Becarios y Tesistas de la Patagonia Austral*, pp. 750-754, ISBN 978-987-3714-88-7, 1ra. Ed Digital 2021.
- [17] Vega, M., Gaetán, G., Martín, A., Saldaño, V. Realidad Aumentada para Mejorar Experiencias de Turismo Cultural. *VI Encuentro de Investigadores, Becarios y Tesistas de la Patagonia Austral*, pp. 787-791, ISBN 978-987-3714-88-7, 1ra. Ed Digital, 2021.
- [18] Martín, A., Gaetán, G., Saldaño, V., Cardozo, C., Vega, M., Sosa, H., Villagra, S., Carrizo, A., Contextos Digitales para Asistencia de Los Ciudadanos: Enfoques de Experiencia de Usuario. *XXIII Workshop de Investigadores en Cs. de la Computación (WICC)*; pp. 421-425. ISBN: 978-987-24611-3-3, <<http://sedici.unlp.edu.ar/handle/10915/120164>>, 2021.
- [19] Vega, M., Gaetán, G., Martín, A., Guías de Experiencia de Usuario para Aplicaciones de Turismo Cultural basadas en Realidad Aumentada. *ICT-UNPA*, Vol. 13(2), pp. 26-43 ISSN: 1852-4516, DOI: <<https://doi.org/10.22305/ict-unpa.v13.n2.812>>, 2021.

Proceso de validación de requerimientos de software

Sonia R. Santana , Leandro Antonelli , Pablo Thomas 

Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos
sonia.santana@uner.edu.ar
Laboratorio de Investigación y Formación en Informática Avanzada (LIFIA), Facultad de
Informática, Universidad Nacional de La Plata
leandro.antonelli@lifia.info.unlp.edu.ar
Instituto de Investigación en Informática (III-LIDI), Facultad de Informática,
Universidad Nacional de La Plata
pthomas@lidi.info.unlp.edu.ar

Resumen

En el marco de la Ingeniería de Requerimientos (RE por sus siglas en inglés Requirements Engineering) la validación de los requerimientos es una tarea fundamental en cualquier proyecto de Ingeniería de Software y debe ser un proceso continuo en el ciclo de vida del desarrollo del sistema. El principal objetivo de la validación de los requerimientos es confirmar que los requerimientos especificados sean representaciones de las necesidades y expectativas de los usuarios [1] [2] [3] y deben ser completos, correctos y consistentes [3] entre otras características.

Los objetivos establecidos en esta línea de investigación permitirán obtener información sobre los problemas asociados al proceso de validación de requerimientos en el ciclo de vida del software: la naturaleza de la información Quién, Qué, Cuándo, Por qué y Cómo validar los requerimientos y analizar comparativamente sus características, necesidades de información y restricciones en base a determinados criterios sobre los enfoques analizados.

Palabras clave: Ingeniería de Requerimientos, validación de requerimientos, enfoques para la ingeniería de requerimientos.

Contexto

El presente trabajo se encuadra en la línea de investigación "Ingeniería de Software", establecida como prioritaria desde la carrera Licenciatura en Sistemas de la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos (UNER). Se adecua, además, a una de las prioridades de la UNER considerando que es un trabajo aplicado a la investigación sobre Tecnologías de la Información y la Comunicación.

Introducción

En la industria del software el objetivo fundamental que se debería alcanzar es obtener productos software de calidad aplicando diferentes métodos y modelos bajo procesos estandarizados en el desarrollo de software, combinando técnicas, estrategias y todos aquellos aspectos que contribuyan a la búsqueda de conseguir calidad en los productos intangibles.

En la actualidad, se observa que los sistemas de software son cada vez más complejos, por lo que las exigencias de calidad en el desarrollo deber ser mayores. Es necesario que el desarrollo de software sea más riguroso para obtener un

producto de alta calidad. Esto busca reducir el número de errores presentes en el desarrollo. La identificación y tratamiento de estos errores en etapas tempranas del proyecto de software es crucial para disminuir los costos de operación y evitar errores en etapas posteriores del desarrollo.

"La parte más difícil de desarrollar un sistema de software es decidir exactamente qué construir. Ninguna otra parte del trabajo conceptual es tan difícil como establecer los requerimientos técnicos detallados... Ninguna otra parte del trabajo paraliza tanto el sistema resultante si se hace mal. Ninguna otra parte es tan difícil de rectificar posteriormente" [4].

Según Bahill, el proceso de validación de requerimientos consiste en primer lugar asegurar que un conjunto de requerimientos son: correctos, completos y consistentes, en segundo lugar si se puede crear un modelo que cumpla con los requerimientos, y por último se pueda construir y probar una solución de software en el mundo real para demostrar que cumple con los requerimientos de las partes interesadas [5].

Según Kof, el 56 por ciento de los errores en los proyectos de software se deben a los errores generados en la fase de requerimientos, como se muestra en la Figura 1. La mitad de estos errores se deben a los requerimientos incompletos y ambiguos y la otra mitad se debe a los requerimientos que se omiten [6].

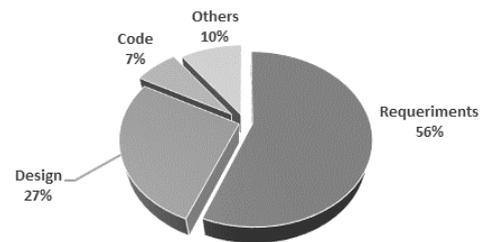


Figura 1. Distribución de errores [6].

Trabajar en el contexto de la validación de requerimientos se está convirtiendo en un desafío para los equipos, clientes y usuarios, existen diferentes causas que imponen problemas de comunicación, control, intercambio de conocimientos, confianza y retrasos en el desarrollo del software [7].

La investigación en el área de ingeniería de requerimientos ha sido reconocida desde mediados de la década del 80 [8].

Diferentes revisiones sistemáticas publicadas demuestran el interés en la investigación empírica de la ER que está en aumento en su conjunto. Sin embargo, el análisis de enfoques para la validación de los requerimientos, a pesar de su reconocida importancia, la investigación empírica es todavía incipiente en la actualidad [9].

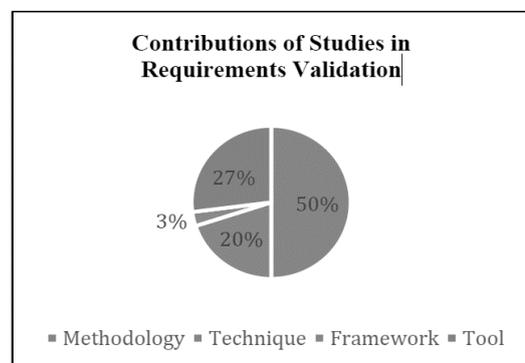


Figura 2. Contribución de estudios en la validación de requerimientos desde el año 2007 hasta el año 2016 [9].

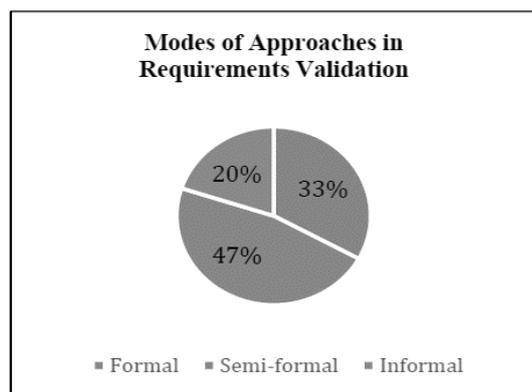


Figura 3. Modos de enfoques en la validación de los requerimientos [9].

El estudio realizado por Moketar establece que muchos enfoques validan los requerimientos funcionales utilizando el enfoque/método semiformal. Para ello, el modelo UML fue utilizado con más frecuencia para la validación de requerimientos. En cuanto a las técnicas, la creación de prototipos fue la favorita, seguida por la simulación, la validación de requerimientos basada en modelos y basada en pruebas. El análisis también encontró que los criterios de calidad más importantes de los requerimientos eran la consistencia, corrección y lo completo [9].

Sourour, en el desarrollo de la metodología para la validación de requerimientos en un entorno colaborativo plantea problemas durante el proceso de validación de requerimientos: ¿Cuál es el resultado de la validación, cómo podemos validar los requerimientos? ¿Qué tipos de procesos de validación son los más adecuados para un proyecto? ¿Cómo garantizar que la solución satisfaga las necesidades de las partes interesadas y de la empresa? ¿Cuál es lo mejor técnica a utilizar en la validación? ¿Cómo ponerse de acuerdo todas las partes interesadas? Los diferentes modos de validación (formal,

semi-formal, informal). Control de los requerimientos funcionales cambiantes, técnicas de negociación insuficientes para la validación, la falta de actividad de validación en RE, la falta de métodos de validación, la falta de expertos analistas, falta de expertos en negocios con un alto nivel analítico y comunicación entre otros [10].

En esta línea de investigación se analizan los problemas asociados al proceso de validación de requerimientos en el ciclo de vida del software: la naturaleza de la información Quién, Qué, Cuándo, Por qué y Cómo validar los requerimientos, específicamente analizando la información de distintos enfoques de la validación de requerimientos.

Líneas de Investigación, Desarrollo e Innovación

- Procesos de Ingeniería Requerimientos
- Enfoques de Validación de Requerimientos
- Técnicas de Validación de Requerimientos
- Rol de los stakeholders en el proceso de Validación de Requerimientos.
- Cualidades de la SRS a evaluar en el proceso de Validación de Requerimientos.
- Técnicas de procesamiento de lenguaje natural.
- Procesos Colaborativos.

Resultados obtenidos/esperados

Los resultados obtenidos / esperados se pueden resumir en:

- Se ha avanzado en una revisión bibliográfica de las principales tendencias de la Validación de Requerimientos del software desde el año 2007 hasta el año 2021. Los

artículos analizados fueron obtenidos desde las fuentes IEEE, Elsevier, Springer y ACM Digital Library. Se han preseleccionado 38 trabajos para posteriormente centralizarse en 4 metodologías [11].

- Se ha avanzado en el análisis comparativo de cuatro metodologías con el fin de identificar las contribuciones al proceso de Validación de Requerimientos [11]:
 - Funciones, componentes, entornos y características.
 - Actividades de planificación.
 - Técnicas de control e indicadores rendimiento.
 - Definición de estándares.
 - Aceptación del cliente/usuario.
 - Dominios de aplicación.
 - Participación cliente/usuario.
 - Etapas del ciclo de vida del desarrollo del software donde validan los requerimientos.
- Avanzar en la capacitación continua de los miembros de la línea de investigación.
- Avanzar en el aprendizaje de procesos de Ingeniería de Requerimientos.
- Avanzar en el aprendizaje de técnicas de procesamiento de lenguaje natural teniendo como finalidad aplicarlas al proceso de Validación de Requerimientos.
- Avanzar en el estudio de las cualidades de la SRS a evaluar en el proceso de Validación de Requerimientos.
- Avanzar en el estudio de procesos colaborativos teniendo como finalidad aplicarlos al proceso de Validación de Requerimientos.

Formación de Recursos Humanos

Este estudio prevé al menos, dos proyectos de Trabajo Final de la carrera Licenciatura en Sistemas y la realización de un trabajo de especialización y una tesis de maestría en la Facultad de Informática de la UNLP.

Referencias

1. P. A. Laplante: Requirements Engineering for Software and Systems, CRC Press (2019).
2. B. H. C. Cheng, J. M. Atlee: Current and Future Research Directions in Requirements Engineering, Design Requirements Engineering A Ten-Year Perspective, Lecture Notes in Business Information Processing, vol. 14, pp. 11–43 (2019).
3. S. L. Pfleeger: Software Engineering – Theory and Practice, Prentice Hall (1998).
4. F. Brooks: The Mythical Man Month: Essays on Software Engineering. Reading, Mass, USA, Addison-Wesley, (1975).
5. A. Terry Bahill, Steven J. Henderson: Requirements development, verification, and validation exhibited in famous failures, Systems Engineering. 8. 1 - 14. 10.1002/sys.20017 (2005).
6. L. Kof, R. Gacitua, M. Rouncefield, P. Sawyer: Ontology and Model Alignment as a Means for Requirements Validation. IEEE Fourth International Conference on Semantic Computing, 2010, pp. 46-51, (2010).
7. P. Loucopoulos, V. Karakostas: System Requirements Engineering, McGraw-Hill, London, ISBN 0-07-707843-8 (1995).

8. A. van Lamsweerde: Requirements Engineering from System Goals to UML Models to Software Specification, 2nd edition. Wiley, (2009).
9. Mokhtar, Nor Aiza, Kamalrudin, Massila, Mohd-Yusof, Mokhtar, Sidek, Safiah: A review on requirements validation for software development, Journal of Theoretical and Applied Information Technology. 96. 3182-3193, (2018).
10. Sourour Maalem, Nacereddine Zarour: Challenge of validation in requirements engineering, Journal of Innovation in Digital Ecosystems, Volume 3, Issue 1, pp 15-21, (2016).
11. Sonia R. Santana, Leandro Antonelli, Pablo Thomas: Evaluación de metodologías para la validación de requerimientos, XXVII Congreso Argentino de Ciencias de la Computación (CACIC), pp. 419-428, ISBN 978 -987-633-574-4, (2021).
12. G. Kotonya, I. Sommerville: Requirements Engineering: Processes and Techniques, John Wiley & Sons, (2004).
13. I. Sommerville, P. Sawyer: Requirements Engineering: A Good Practice Guide, Wiley, (2006).
14. S. L. Pfleeger: Software Engineering – Theory and Practice, Prentice Hall (1998).
15. Pressman S.R.: Ingeniería del software. Un enfoque práctico, Séptima edición, México D.F., Mc Graw Hill, (2010).

Aplicaciones Móviles 3D con Realidad Virtual y Realidad Aumentada

Pablo Thomas , Federico Cristina , Sebastián Dapoto , Patricia Pesado 

Instituto de Investigación en Informática LIDI (III-LIDI)

Facultad de Informática – Universidad Nacional de La Plata

50 y 120 La Plata Buenos Aires

Centro Asociado a la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

{pthomas, fcristina, sdapoto, ppesado}@lidi.info.unlp.edu.ar

▪ Resumen

Se presenta una línea de investigación y desarrollo, que tiene por objeto estudiar temas relacionados con aspectos de Ingeniería de Software, orientados al desarrollo de aplicaciones móviles tridimensionales (3D) sobre diversas plataformas. En particular, se pone el foco en el desarrollo de aplicaciones 3D inmersivas, a través del uso de Realidad Virtual (RV) y Realidad Aumentada (RA). También se realizan desarrollos de aplicaciones móviles 3D que utilizan sensores inteligentes.

Palabras claves: Dispositivos Móviles - Aplicaciones 3D - Aplicaciones Multiplataforma - M-Learning – Realidad Virtual – Realidad Aumentada – Sensores Inteligentes

▪ Contexto

Esta línea de Investigación forma parte del proyecto “*Metodologías, técnicas y herramientas de ingeniería de software en escenarios híbridos. Mejora de proceso.*”, en particular del subproyecto “*Ingeniería de Software para escenarios híbridos*”, del Instituto de Investigación en Informática LIDI de la Facultad de Informática, acreditado por el Ministerio de Educación de la Nación.

Existe una importante cooperación con Universidades de Argentina y se está trabajando con Universidades de Europa en proyectos financiados por el Ministerio de Ciencia y Tecnología de España y la AECID.

Se participa en iniciativas como el Programa IberoTIC de intercambio de Profesores y

Alumnos de Doctorado en el área de Informática.

Por otra parte, se tiene financiamiento de Telefónica de Argentina en Becas de grado y posgrado.

▪ Introducción

En la actualidad los dispositivos móviles permiten ejecutar aplicaciones complejas y con exigentes requerimientos de hardware. Debido a esto, existen cada vez más alternativas de motores de juego que permiten desarrollar aplicaciones tridimensionales para dispositivos móviles.

La RV es una simulación interactiva por computadora en la cual se sustituye el mundo real con información sensorial que recibe el usuario. La RV permite generar entornos inmersivos donde el usuario puede interactuar con representaciones virtuales de objetos, que de otro modo sería difícil o imposible de acceder [1].

La RA es el conjunto de tecnologías que permiten que el usuario visualice el mundo real con información virtual añadida, mediante un dispositivo tecnológico. Los elementos físicos reales y tangibles se combinan con elementos virtuales. La RA es interactiva y en tiempo real [2].

El ámbito educativo debe adaptarse a los cambios y nuevas formas de aprendizaje. M-learning (mobile learning) plantea métodos modernos de apoyo al proceso de aprendizaje mediante el uso de dispositivos móviles. Las

aplicaciones móviles 3D son una herramienta ideal para acercar a los alumnos [3] [4] [5].

Los sistemas de RA pueden transformar la forma de aprender y trabajar [6]. Mediante la utilización de gafas de RA y sensores inteligentes, es posible, por ejemplo, asistir a un técnico que realiza tareas de mantenimiento y deba realizar procedimientos con las manos libres.

▪ Líneas de Investigación y Desarrollo

- Metodologías y Técnicas de la Ingeniería de Software y su aplicación en el desarrollo de software para dispositivos móviles.
- Aplicaciones Móviles 3D Multiplataforma
- Mobile Learning
- Frameworks para el desarrollo de Aplicaciones Móviles 3D
- Realidad Virtual en aplicaciones móviles 3D
- Realidad Aumentada en aplicaciones móviles 3D
- Aplicaciones móviles 3D que utilizan sensores inteligentes

▪ Resultados esperados/obtenidos

Los resultados esperados/obtenidos se pueden resumir en:

- Avanzar en la capacitación continua de los miembros de la línea de investigación.
- Avanzar en el aprendizaje de frameworks o motores de juego que permiten desarrollar aplicaciones 3D multiplataforma, particularmente para dispositivos móviles [7] [8] [9] [10].
- Avanzar en el desarrollo de aplicaciones educativas, teniendo como finalidad enriquecer las experiencias interactivas y motivar el aprendizaje mediante su uso.
- Avanzar en el desarrollo de aplicaciones móviles 3D relacionadas con sensores inteligentes.

- Avanzar en el desarrollo de aplicaciones móviles 3D con Realidad Virtual y Realidad Aumentada.
- Se ha desarrollado la aplicación móvil R-Info3D [3]. Se trata de una herramienta de aprendizaje sencilla de los conceptos básicos para la construcción de algoritmos en la Facultad de Informática. El intérprete de código permite el uso de variables, estructuras de control, operadores matemáticos, múltiples robots y paralelismo, entre otros. Además, la aplicación puede ser utilizada con lentes de RV, permitiendo una inmersión completa en el escenario virtual. Figuras 1 y 2.
- Se ha desarrollado una aplicación móvil 3D de domótica denominada 3D-Domotic [11]. La aplicación facilita la interacción con los objetos conectados a una vivienda, mediante controles visuales, cómodos e intuitivos. El servidor puede estar en una red local o en la nube, permitiendo el control remoto de los dispositivos. Figuras 3 y 4.



Figura 1. La aplicación móvil R-Info3D.

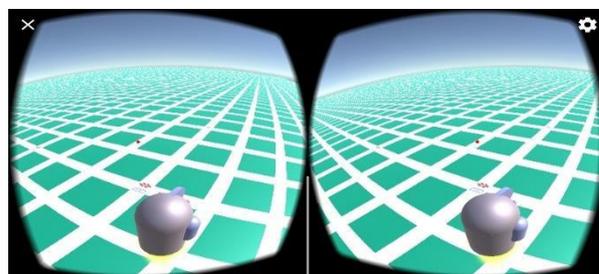


Figura 2. La aplicación móvil R-Info3D utilizado con lentes de RV.

- Se está avanzando en el desarrollo de un prototipo móvil 3D de asistencia técnica por medio de realidad aumentada y sensores inteligentes.

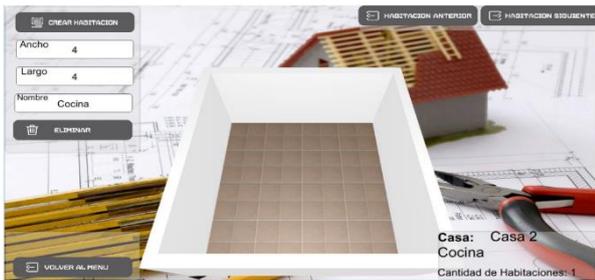


Figura 3. Aplicación móvil 3D de domótica. Creación de los ambientes que componen la casa que se desea controlar.



Figura 4. Aplicación móvil 3D de domótica. Control visual de los dispositivos.

Formación de Recursos Humanos

Los integrantes de esta línea de investigación dirigen Tesinas de Grado y Tesis de Postgrado en la Facultad de Informática, y Becarios III-LIDI en temas relacionados con el proyecto. Además, participan en el dictado de asignaturas/cursos de grado y postgrado de la Facultad de Informática de la UNLP.

Referencias

- Linowes J. "Unity Virtual Reality Projects". 2015. ISBN-13: 978-1783988556.
- R. Silva, J.C. de Oliveira, G. Giraldi. "Introduction to augmented reality". 2003.
- Federico Cristina, Sebastián Dapoto, Pablo Thomas, Patricia Pesado. Capítulo de Libro: "3D Mobile Prototype for Basic Algorithms Learning". Libro: "Computer Science & Technology Series - XXI Argentine Congress Of Computer Science. Selected Papers" (300 páginas). EDULP. ISBN: 978-987-4127-00-6, páginas 239-247. Año 2016.
- Kantel E., Tovar G., Serrano A. "Diseño de un Entorno Colaborativo Móvil para Apoyo al Aprendizaje a través de Dispositivos Móviles de Tercera Generación." IEEE-RITA 5, no. 4 (2010): 146-151.
- Rosa Paredes, J. Alfredo Sánchez, Liliana Rojas, Daniel Strazzulla, Ronel Martínez-Teutle. "Interacting with 3D Learning Objects". 2009 LA Web Congress. ISBN: 978-0-7695-3856-3/09.
- X. Pan, X. Sun, H. Wang, S. Gao, N. Wang and Z. Lin, "Application of an assistant teaching system based on mobile augmented reality (AR) for course design of mechanical manufacturing process," 2017 IEEE 9th International Conference on Engineering Education (ICEED), 2017, pp. 192-196, doi: 10.1109/ICEED.2017.8251191.
- Unity 3D Homepage: <https://unity3d.com/>.
- Unreal Engine Homepage. <https://www.unrealengine.com/>.
- CryEngine Homepage. <https://www.cryengine.com/>
- Godot Engine Homepage. <https://godotengine.org>
- "3D-Domotic: A 3D Mobile Application for Domotic Control". X Jornadas de Cloud Computing, Big Data & Emerging Topics JCC-BD&ET 2021. Autores: Sebastián Dapoto, Diego Encinas, Federico Cristina, Cristian Iglesias, Federico Arias, Pablo Thomas, Patricia Pesado. ISBN: 978-3-030-84825-5, vol. 14444, págs. 165-176, doi. 10.1007/978-3-030-84825-5_12, 2021.

Reuso de información en comunidades virtuales

Gabriela Aranda, Valeria Zoratto, Nadina Martinez Carod, Romina Schroeder
Andrés Flores, Natalia Baeza, Lucas Cavaliere, Sandra Lucero

Grupo de Investigación en Ingeniería de Software del Comahue (GIISCO)
<http://giisco.uncoma.edu.ar>

Facultad de Informática. Universidad Nacional del Comahue
Buenos Aires 1400, (8300) Neuquén

Contacto: {gabriela.aranda, vzoratto, nadina.martinez}@fi.uncoma.edu.ar

RESUMEN

Con el comienzo del nuevo siglo, la Web 2.0 permitió que cualquier usuario con conexión a internet y sin conocimientos especiales, pueda publicar contenido propio mediante herramientas colaborativas como blogs, wikis, y foros de discusión. Al mismo tiempo surgieron plataformas para creación de redes sociales de distinto tipo o con objetivos diferentes, concretamente se destaca el trabajo distribuido colaborativo, el intercambio de conocimiento técnico, académico, científico y lo social. Estas comunidades que fueron formándose a partir de la participación online recibieron el nombre de comunidades virtuales.

Durante los últimos años hubo una evolución hacia la Web 3.0 la cual se ocupa de, por un lado, facilitar la accesibilidad a la información sin depender del dispositivo utilizado, y por otro lado, el modo que las personas interactúan con ella para conseguir los resultados que desean. Además, se enfoca en que la información sea compartida de una forma inteligible, que sea útil para los usuarios que recurren a ella y a las necesidades especiales de sus usuarios de acuerdo a las circunstancias. Así, nuestro proyecto se enfocará en proveer modelos de calidad para sistemas software de recuperación, análisis, clasificación y reuso de la información proveniente de comunidades virtuales en la Web.

Palabras clave

Recuperación de Información, calidad de datos, comunidades virtuales, participación ciudadana.

CONTEXTO

Esta línea de investigación forma parte del proyecto de investigación “Reuso de información en comunidades virtuales”, de la Universidad Nacional del Comahue, en estado de evaluación, con período de vigencia 2022-2025. Dicha línea extiende y avanza sobre temas desarrollados por el equipo de investigadores principales en proyectos anteriores, pertenecientes al Programa “Desarrollo de Software basado en Reúso”, de la Universidad Nacional del Comahue, llevado a cabo entre 2013-2021. El nuevo proyecto continúa las líneas de investigación enfocadas en la recuperación de información disponible en foros de discusión y abarca nuevas tecnologías para soporte a comunidades virtuales con una mirada orientada a la participación ciudadana y la toma de decisiones basada en la opinión pública.

1. INTRODUCCIÓN

El término “comunidades virtuales” fue establecido en 1993 por Howard Rheingold para referirse a grupos sociales que emergen a partir de la interacción de personas en

espacios públicos de internet [1]. Dentro de esta definición se enmarcan distintos tipos de entornos colaborativos como las Comunidades de Preguntas y Respuestas (CQA, del inglés *Community Question Answering*), las plataformas para redes sociales (como Facebook, Twitter o Instagram), entre otras.

Particularmente, las CQA ofrecen a los usuarios la posibilidad de buscar y compartir conocimiento a partir de la formulación de preguntas, respondiendo las mismas y/o comentando. Entre las CQA se encuentran los foros de discusión, los cuales permiten el debate acerca de un tema en particular en forma de hilos de mensajes, teniendo una estructura similar aún entre distintas plataformas existentes. Por ejemplo, sitios como Quora (que permite hacer preguntas de todo tipo) y otros de dominios más específicos, como StackOverflow (para programación) o Mathematics StackExchange (para matemáticas), son sitios utilizados diariamente por millones de usuarios para encontrar respuesta a preguntas complejas, subjetivas o dependientes del contexto [2, 3]. Con el tiempo, creció considerablemente la cantidad de información acumulada en dichos sitios, por lo que analizar y reutilizar dicha información es algo “deseable y valioso” [4, 5]. Así, el reuso e integración de información (IRI, del inglés *Information Reuse and Integration*) juega un rol esencial ya que se enfoca en la captura, representación, mantenimiento, integración, validación y extrapolación de información, que luego puede ser aplicada para mejorar la toma de decisiones en varios dominios de aplicación [6].

Con referencia a esto último, existen varias investigaciones enfocadas en analizar, evaluar y extraer la información proveniente de CQA. Por ejemplo, Le et al. [7] y Amancio et al. [8] analizan los hilos de discusión para determinar cómo se constituye la calidad de las respuestas. En cambio, Neshati [3] evalúa la calidad del contenido desde la perspectiva del resultado de la votación. Por otro lado,

algunos autores consideran que todo lo necesario para evaluar la calidad de una respuesta está en su contenido, es decir, tienen en cuenta sólo las características textuales de los hilos [9, 10, 11], mientras que otros combinan tanto las características textuales como la red de usuarios [12, 13]. Otros trabajos se enfocan en la calidad del contenido en base al nivel de experticia del usuario que responde, pero dado que no todas las CQA mantienen información de la reputación de un usuario como un puntaje o categoría asociada, una línea de investigación relevante es la de estimar su experticia basado en las evidencias disponibles, como la calidad de sus intervenciones y su interrelación con otros usuarios [14]. Es decir que tanto el análisis del contenido textual de los mensajes como el estudio de las redes sociales subyacentes son temas muy relacionados a este tipo de investigación.

Respecto al análisis de dichas redes sociales se requiere de un conjunto de métodos y técnicas en los que se mezclan teorías sociológicas y matemáticas. Para analizar las estructuras sociales de estas redes se utilizan los conceptos, el vocabulario y operaciones de la teoría de grafos que permiten probar teoremas sobre los grafos que las modelan, así como deducir y someter determinados enunciados a tests [15]. Algunos de los trabajos más actuales en esta área están enfocados en el estudio de información proveniente de redes sociales como Twitter [16], pero dado que cada red tiene intereses y características únicas, es pertinente evaluar el comportamiento de las personas en otros tipos de plataformas colaborativas; por ejemplo para detectar usuarios que emiten información irrelevante, los que son líderes de opinión (vistos como fuentes de conocimiento dentro de su comunidad) o los denominados *boundary spanner* (usuarios que permiten comunicación entre distintas comunidades) [17].

Como se ha mencionado, el gran volumen de información generado por las CQA y las redes sociales existentes en la Web es

propicio para estudiar y definir técnicas para reuso e integración de información [6], por lo que este proyecto se enfocará en definir técnicas para captura de información, análisis de contenido, detección de redes sociales y clasificación de perfiles de usuarios utilizando para su evaluación corpus de comunidades virtuales existentes como StackExchange¹, que es una red de webs para CQA que cuenta con más de 50 foros de discusión de diferentes temáticas, 14 millones de usuarios, 21 millones de preguntas y 31 millones de respuestas, cuyas bases de datos son accesibles de forma libre para investigación². Este conocimiento puede ser aplicado en comunidades virtuales más específicas, como por ejemplo, las conformadas a partir de plataformas para la participación ciudadana. Este tipo de plataformas han surgido en los últimos años a partir de los desafíos que enfrentan las ciudades para garantizar la calidad de vida de sus habitantes y mejorar los procesos de toma de decisiones incorporando la opinión pública. Por ello, la participación ciudadana es una herramienta que mejora la gobernanza local y la toma de decisiones, que busca ser una forma directa de conocer las necesidades, demandas e ideas de los individuos que la componen [18]. En muchos casos la toma de decisión se realiza a partir de la opinión de la ciudadanía obtenida mediante herramientas colaborativas, como sitios web y/o aplicaciones móviles, que permiten conocer a corto plazo los efectos que pueden tener los cambios realizados, por ejemplo, en el espacio urbano. Dado que dichas tecnologías se basan en recuperar opiniones de ciudadanos, surgen temáticas de análisis y evaluación que, muchas veces son comunes a otras comunidades virtuales como las que hemos mencionado anteriormente. Con este objetivo, como parte de este proyecto se evaluarán productos existentes para participación ciudadana (como *decidim*³,

*CONSUL*⁴ y *WeLive*⁵ [19, 20]) y se trabajará en su adaptación para aplicarlos en el ámbito de barrios de la ciudad de Neuquén. Además, se planea hacer uso de las técnicas y herramientas elaboradas como parte de otras líneas de investigación, para hacer aportes al reuso de información como soporte para la toma de decisiones basada en la opinión de los ciudadanos.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El proyecto de investigación actual se denomina “Reuso de información en comunidades virtuales” y su objetivo principal es definir cuáles son las mejores técnicas y algoritmos de recomendación para la asistencia inteligente a usuarios de comunidades virtuales en la búsqueda de información relevante

Este proyecto está desarrollado por integrantes del Grupo de Ingeniería de Software de la Universidad Nacional del Comahue, (GIISCo), formado por docentes y estudiantes de la Facultad de Informática de la Universidad Nacional del Comahue, junto con asesoría y colaboración de otras universidades. En particular, este proyecto es desarrollado en colaboración con la Facultad de Ciencias Exactas de la Universidad Nacional del Centro de la Provincia de Buenos Aires. El proyecto también involucra a docentes pertenecientes a otras áreas de la Facultad, como Programación, Ingeniería en Computación, Ingeniería en Sistemas y Teoría de la Computación, lo que permite abordar la investigación desde ópticas diferentes, enriqueciendo el desarrollo con un trabajo conjunto y colaborativo.

3. RESULTADOS OBTENIDOS/ESPERADOS

A partir del 2013 hemos realizado investigación en temas relacionados a

¹ <https://stackexchange.com/>

² <https://archive.org/download/stackexchange>

³ <https://decidim.org/es/>

⁴ <https://consulproject.org/en/>

⁵ <https://welve.eu/>

comunidades virtuales de programadores, proponiendo un modelo de calidad para foros de discusión técnicos y determinando criterios para la calidad de información contenida en un hilo de discusión [22], que fueron validados mediante encuestas y se formularon variaciones en la parametrización para mejorar los resultados obtenidos [23].

También se trabajó en el procesamiento del texto de hilos de discusión de foros técnicos, implementando una herramienta ad-hoc (basada en Lucene) para recuperación de información y análisis según un conjunto propio de medidas de calidad para proponer un ranking de soluciones posibles para una pregunta [24]. Luego, se mejoró dicha herramienta incluyendo sinónimos con la base de datos léxica WordNet y el analizador morfológico Stanford POS Tagger para identificar el rol de las palabras en el contexto que eran utilizadas [25].

Además, se avanzó en la clasificación de roles de usuarios activos de un foro, para determinar la jerarquía de roles determinados por el nivel de conocimiento de los participantes en los hilos de discusión de acuerdo a los posts realizados por dichos usuarios [26].

Utilizando dichos conocimientos obtenidos por este grupo de trabajo en estudios anteriores, se espera aplicarlos en el análisis de información proveniente de comunidades virtuales creadas ad-hoc para los barrios de nuestra ciudad y aportar a la toma de decisiones basada en opinión pública.

4. FORMACIÓN DE RECURSOS HUMANOS

El proyecto se encuentra conformado por docentes de diferentes áreas debido a su naturaleza multidisciplinaria. Las personas que forman parte del proyecto, tanto como colaboradores, asesores o integrantes son:

- Tres docentes investigadores del Departamento de Programación e

Ingeniería de Sistemas, con dedicación exclusiva, con Doctorado en Informática.

- Una docente investigadora del Departamento de Programación con beca del CONICET para realización de doctorado.
- Tres docentes con dedicación simple, pertenecientes a los Departamentos de Programación, Ingeniería de Sistemas e Ingeniería en Computación.
- Una profesora adjunta, asesora local, con dedicación exclusiva, del Departamento de Teoría de la Computación.
- Una docente investigadora externa, perteneciente al Instituto Superior de Ingeniería del Software (ISISTAN) de la Universidad Nacional del Centro de la Provincia de Buenos Aires (UNCPBA), con experiencia en Sistemas de Recomendación y Recuperación de Información. Doctora en Ciencias de la Computación.
- Tres estudiantes de la carrera de Licenciatura en Ciencias de la Computación que realizan sus tesis dentro del proyecto.

De esta manera, se van incorporando actividades para extender líneas de investigación al proyecto inicial con nuevos enfoques.

5. BIBLIOGRAFÍA

- [1] H. Rheingold. *The Virtual Community*, revised edition: Homesteading on the Electronic Frontier. MIT press, 2000.
- [2] I. Srba and M. Bielikova. A comprehensive survey and classification of approaches for community question answering. *ACM Trans. Web*, 10(3), 2016.
- [3] M. Neshati. On early detection of high voted qa on stack overflow. *Information Processing Management*, 53(4):780–798, 2017.
- [4] G. Cong, L. Wang, C. Lin, Y. Song, and Y. Sun. Finding Question-answer Pairs from Online Forums. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '08*, pages 467–474, New York, NY, USA, 2008. ACM.

- [5] S. Gottipati, D. Lo, and J. Jiang. Finding relevant answers in software forums. In 26th IEEE/ACM International Conference on Automated Software Engineering (ASE 2011), Lawrence, KS, USA, November 6-10, 2011, pages 323–332, 2011.
- [6] M. Day, C. Ong, and W. Hsu. An analysis of research on information reuse and integration (2003-2008). *International Transactions on Systems Science and Applications*, 6(2):146–157, 2010.
- [7] L.T. Le, C. Shah, and E. Choi. Evaluating the Quality of Educational Answers in Community Question-Answering. In Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries, JCDL '16, pages 129–138, New York, NY, USA, 2016. Association for Computing Machinery.
- [8] L. Amancio, C. Dorneles, and D. Dalip. Recency and quality-based ranking question in CQAs: A stack overflow case study. *Information Processing Management*, 58(4):102552, 2021.
- [9] G. Gkotsis, K. Stepanyan, C. Pedrinaci, J. Domingue, and M. Liakata. It's all in the content: state of the art best answer prediction based on discretisation of shallow linguistic features. In Proceedings of the 2014 ACM conference on Web science, pages 202–210, 2014.
- [10] G. Burel, P. Mulholland, and H. Alani. Structural Normalisation Methods for Improving Best Answer Identification in Question Answering Communities. In Proceedings of the 25th International Conference Companion on World Wide Web, pages 673–678, 2016.
- [11] Y. Pérez-Guadarramas, A. Rodríguez-Blanco, A. S. Cuevas, W. Hojas-Mazo, y J. A. Olivas. Combinando patrones léxico-sintácticos y análisis de tópicos para la extracción automática de frases relevantes en textos. *Procesamiento del Lenguaje Natural*, (59):39–46, 2017.
- [12] D. Kundu and D. Prasad Mandal. Formulation of a hybrid expertise retrieval system in community question answering services. *Applied Intelligence*, 49(2):463–477, 2019.
- [13] H. Fu and S. Oh. Quality assessment of answers with user identified criteria and data-driven features in social qa. *Information Processing Management*, 56(1):14–28, 2019.
- [14] M. Neshati, Z. Fallahnejad, and H. Beigy. On dynamicity of expert finding in community question answering. *Information Processing Management*, 53(5):1026–1042, 2017.
- [15] L. Sanz-Menéndez. Análisis de redes sociales: O cómo representar las estructuras sociales subyacentes. *Apuntes de Ciencia y Tecnología*, 7:21–29, 06 2003.
- [16] R. Olivares, F. Muñoz, and F. Riquelme. A multiobjective linear threshold influence spread model solved by swarm intelligence-based methods. *Knowledge-Based Systems*, 212:106623, 2021.
- [17] P. Matous and P. Wang. External exposure, boundary-spanning, and opinion leadership in remote communities: A network experiment. *Soc. Networks*, 56:10–22, 2019.
- [18] D. García Castro, V. De Elizagarate Gutierrez, J. Kazak, S. Szewranski, I. Kaczmarek, and T. Wang. Nuevos desafíos para el perfeccionamiento de los procesos de participación ciudadana en la gestión urbana. retos para la innovación social. *Management Letters/Cuadernos de Gestión*, 20(1):41–64, 2020.
- [19] I. Peña-López. Shifting participation into sovereignty: the case of decidim.barcelona. 03 2019.
- [20] M. X. Rivera Rásury. Desarrollo de una herramienta de soporte metodológico a los procesos de e-participación. Master, Departamento de Sistemas Informáticos y Computación Universitat Politècnica de Valencia, Valencia, España, 2018.
- [21] J. Levy Moreno and H. H. Jennings. “Statistics of Social Configurations.” *Sociometry*, vol. 1, no. 3/4, pp. 342–374. *JSTOR*, 1938.
- [22] G. Aranda, N. Martínez Carod, S. Roger, P. Faraci, A. Cechich, V. Zoratto. Una herramienta para el análisis de hilos de discusión técnicos. *CACIC 2014*, Buenos Aires, pp.803-812, 2014.
- [23] G. Aranda, V. Zoratto, N. Martínez Carod, S. Roger, F. Otermin, A. Cechich. Clasificación de contenido de hilos de discusión mediante análisis sintáctico y morfológico. *CICCSI 2018*. ISBN 9789874568366. Mendoza, 2018, pp. 35-44.
- [24] V. Zoratto, G. Aranda, S. Roger, A. Cechich. Analyzing Discussion Forums Threads About Java Programming Language Usage, *Electronic Journal of SADIO*, ISSN 1514-6774, 2016.
- [25] V. Zoratto, G. Aranda, N. Martinez Carod, F. Otermin. Evaluación de estrategias para clasificar hilos de foros de discusión según su contenido, *ASSE-JAIIO 2021*, Argentine Symposium on Software Engineering, Argentina, 2021.
- [26] N. Martínez Carod, G. Aranda, V. Zoratto, C. Murray (2019), Una propuesta para clasificación de roles de usuarios en foros de discusión técnicos. *CACIC 2019*, Argentina, 2019.

DESARROLLO DE ESTUDIOS EMPÍRICOS EN INGENIERÍA DE SOFTWARE

Emanuel Irrazabal; Rubén Alfredo Bernal; Juan Andrés Carruthers; Andrea Lezcano Airaldi; Yanina Medina; Iván Sambrana; Celeste Ojeda Rodríguez

Facultad de Ciencias Exactas y Naturales y Agrimensura.
Universidad Nacional del Nordeste
{eirrazabal, jacarruthers}@exa.unne.edu.ar

RESUMEN

Este es el primer año del proyecto F01-2021; una derivación de los proyectos F07-2009, F10-2013 y F018-2017, que desarrollaron modelos, métodos y herramientas para la calidad del software. Este proyecto se enfoca en el diseño y desarrollo de estudios empíricos de calidad de software, como un insumo para la Ingeniería de Software basada en Evidencia.

La línea principal de investigación es la construcción de un catálogo de proyectos software de calidad. Se atiende la necesidad de los grupos de investigación para obtener muestras curadas de proyectos imprescindibles para la generación de resultados confiables y generalizables en la experimentación de estudios empíricos de la calidad de software, proporcionando los insumos y procedimientos necesarios para conseguirlo de manera efectiva.

Otras líneas secundarias de investigación son el desarrollo de estudios empíricos para la toma de decisiones en grupo y el desarrollo de estimaciones por juicio experto mediadas a partir de tecnologías móviles.

Palabras clave: calidad de software, estudios empíricos, proyecto fuente abierta.

CONTEXTO

Las líneas de Investigación y Desarrollo presentada en este trabajo corresponden al

proyecto PI-21F01 “Desarrollo de estudios empíricos en Ingeniería del Software”, acreditado por la Secretaría de Ciencia y Técnica de la Universidad Nacional del Nordeste (UNNE) para el periodo 2022-2025, y a la beca interna doctoral de CONICET otorgada por RESOL-2021-154-APN-DIR#CONICET para el período 2021-2025.

1. INTRODUCCIÓN

En la Ingeniería del Software (IS) se trabaja con la construcción de aplicaciones software multi-versión [1]. Por lo tanto, muchas de las actividades asociadas con una aplicación software provocan revisiones para mejorar la funcionalidad o para corregir errores, especialmente en las metodologías ágiles [2].

En el desarrollo software, la calidad puede estudiarse desde el punto de vista de: i) la calidad del proceso de desarrollo software, y ii) la calidad del código fuente [3]. En este último caso es necesario obtener métodos empíricos para demostrar la calidad del software [4] y utilizar evidencia directamente relacionada con el producto software resultante a partir de métricas e indicadores que se vinculen directamente con la calidad [5]. Sin embargo, esto no siempre es sencillo de realizar, ya que las conclusiones generales de los estudios empíricos en IS a menudo dependen de una gran cantidad de variables [6].

El uso masivo de repositorios para el código fuente (por ej., SourceForge, GitHub o Maven) le ha otorgado a los investigadores e ingenieros de software el

acceso a millones de proyectos y, por lo tanto, datos para el desarrollo de estudios empíricos [7]-[9]. No obstante, la proporción de ruido en una muestra aleatoria tomada de repositorios podría sesgar el estudio, y puede llevar a los investigadores a conclusiones poco realistas, potencialmente inexactas [10].

Esto se contrapone con la condición de reproductibilidad y generalidad de los resultados empíricos, tal y como lo indica el énfasis actual en la IS basada en evidencia. En particular, la reproductibilidad es una condición necesaria, no solo en publicaciones en revistas o conferencias de prestigio de la disciplina, sino también para las empresas de desarrollo de software que a menudo desean analizar la evolución de sus propios proyectos o como patrón comparativo en auditorías de software. En este contexto, una práctica para demostrar la efectividad de las métricas como predictores de las características de calidad del software es la construcción de los denominados corpus o catálogos de proyectos [11].

Los catálogos de proyectos son un insumo para los grupos de trabajo y sirven como mecanismo de comparación para distintos tipos de experimentos. Existen varios ejemplos en la literatura, como los realizados por Barone y R. Sennrich [12], Allamanis y Sutton [13] o Keivanloo [14] y se diferencian por la cantidad, calidad de proyectos que lo componen; como también los criterios y métodos para agruparlos.

Un catálogo popular en IS es el Qualitas corpus [11], cuya última versión es de 2013. En general, se observa que no existe información actualizada en la mayoría de los catálogos, lo que hace necesario revisar los proyectos y sus versiones, y generar nuevas métricas sobre ellos. Una alternativa interesante para ello es la utilización de herramientas de código abierto, que han demostrado ser de gran utilidad en entornos de trabajo profesionales, tanto tradicionales [15] como aplicadas a nuevas técnicas de desarrollo [16].

Por lo tanto, la línea de trabajo del proyecto tiene que ver con las características de calidad del código fuente del software.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

A continuación, se describe la línea de investigación y desarrollo:

Curaduría de proyectos

Se están estudiando los aspectos metodológicos y criterios considerados por la comunidad científica para conformar los catálogos de proyectos. Para esta etapa de documentación se utilizarán dos métodos de investigación. Por un lado, el método de revisiones sistemáticas [17] para identificar, evaluar, e interpretar toda la información relativa a un tema de investigación en particular, de un modo sistemático y replicable. Como segundo método de documentación se utilizarán las encuestas [18], para recopilar información de los grupos de investigación y los equipos de trabajo de las empresas privadas.

El siguiente paso consiste en la creación de un modelo de procedimientos para la construcción, mantenimiento y curaduría de un cuerpo de proyectos software y sus métricas de calidad de producto. Teniendo como fin proveer una estrategia para la construcción de estudios empíricos en IS que brinde una mejor reproductibilidad, consistencia experimental, y flexibilidad para una evolución gestionada del cuerpo de proyectos en el tiempo. Y, finalmente, se propone implementar el modelo en un ambiente real de trabajo.

3. RESULTADOS OBTENIDOS/ESPERADOS

En el marco de este proyecto y respecto de la línea de curaduría de proyectos y de acuerdo con lo desarrollado en el año 2.021 en el marco del proyecto de

investigación anterior se lograron los siguientes objetivos:

1. Relevamiento de información a considerar en la selección de proyectos Software y metadatos utilizados en el campo de estudio.

Se realizaron dos mapeos sistemáticos, uno enfocado en reunir evidencias sobre estudios empíricos en IS realizados con colecciones de proyectos, y el otro en el uso de colecciones existentes en la literatura. El objetivo en ambos casos es identificar las características de los proyectos Software, tipos de metadatos consumidos y las distintas herramientas usadas para recolectar los metadatos de las colecciones. En total se revisaron 5.000 artículos científicos de las principales revistas y congresos sobre la temática.

2. Desarrollo de herramientas para dar soporte en la generación de métricas obtenidas del análisis estático del código fuente.

Sonar Exporting Tool (SET), aplicación web para extraer las métricas de la plataforma Sonar Cloud (SC) en formatos de datos consumibles y difundir el código fuente de los proyectos software analizados. SET permite exportar medidas de métricas de SC en los formatos de datos csv, json y xml; la actualización automática y manual de la base de datos.

Sonar Juploader, aplicación java que permite analizar automáticamente un lote de proyectos con SC por medio del cliente SonarScanner. La aplicación da soporte a la gestión y uso de organizaciones de SC, pre-configuración de proyectos, análisis de proyectos, y visualización de reportes [19].

3. Se realizó un experimento en donde se evaluaron tres modelos predictivos para la detección temprana de defectos en proyectos software [20]. Los tres clasificadores fueron entrenados con tres conjuntos de datos conformados por valores de métricas orientadas a objetos obtenidas del análisis estático del código de cinco sistemas Java. Para el entrenamiento de cada modelo se utiliza el método de ensamble validación y votación.

En este sentido, también se está trabajando en la construcción de un curso que facilite a alumnos de posgrado conceptos esenciales para la experimentación en IS, como también otras áreas del conocimiento. Desde la planificación y diseño de un experimento, hasta la utilización de herramientas estadísticas y visuales para reportarlos de manera efectiva.

4. FORMACIÓN DE RECURSOS HUMANOS

En esta línea de trabajo del Grupo de Investigación sobre Calidad de Software (GICS) están involucrados 3 docentes investigadores, dos becarios internos doctorales de CONICET, un becario de investigación de pregrado y un profesional licenciado en sistemas de información.

5. REFERENCIAS

- [1] Parnas, D. L. (2001). Some software engineering principles. Hoffman, D. M. y Weiss, D. M. (Ed), Software fundamentals: collected papers by David L. Parnas (pp. 257–266). Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.
- [2] Irrazabal, E., Vásquez, F., Díaz, R., & Garzás, J. (2011). Applying

- ISO/IEC 12207: 2008 with SCRUM and Agile methods. In International Conference on Software Process Improvement and Capability Determination (pp. 169-180). Springer, Berlin, Heidelberg.
- [3] Lehman, M. M. (1996). Laws of software evolution revisited. En Montangero C. (Ed), *Laws of Software Evolution Revisited* (pp. 108-124). Berlin, Germany: Springer-Verlag.
- [4] Kitchenham, B. y Pfleeger, S. L. (1996). Software quality: the elusive target. *IEEE Software*, 13(1), 12-21.
- [5] Garvin, D. A. (1984). What Does "Product Quality" Really Mean?. *MIT Sloan Management Review*, 25-43.
- [6] Basili, V. R., Shull, F., & Lanubile, F. (1999). Building knowledge through families of experiments. *IEEE Transactions on Software Engineering*, 25(4), 456-473.
- [7] Vidal, S. A., Bergel, A., Marcos, C., & Díaz-Pace, J. A. (2016). Understanding and addressing exhibitionism in Java empirical research about method accessibility. *Empirical Software Engineering*, 21(2), 483-516.
- [8] Vidal, S., Bergel, A., Díaz-Pace, J. A., & Marcos, C. (2016). Over-exposed classes in Java: An empirical study. *Computer Languages, Systems & Structures*, 46, 1-19.
- [9] Vázquez, H. C., Bergel, A., Vidal, S., Pace, J. D., & Marcos, C. (2019). Slimming javascript applications: An approach for removing unused functions from javascript libraries. *Information and software technology*, 107, 18-29.
- [10] Munaiah, N., Kroh, S., Cabrey, C., & Nagappan, M. (2017). Curating github for engineered software projects. *Empirical Software Engineering*, 22(6), 3219-3253.
- [11] Tempero, E., Anslow, C., Dietrich, J., Han, T., Li, J., Lumpe, M., Melton, H y Noble, J. (2010). The Qualitas Corpus: A curated collection of Java code for empirical studies. *IEEE*, 336-345.
- [12] Barone, A. V. M., y Sennrich, R. (2017). A parallel corpus of Python functions and documentation strings for automated code documentation and code generation. *arXiv preprint arXiv:1707.02275*.
- [13] Allamanis, M., y Sutton, C. (2013). Mining source code repositories at massive scale using language modeling. *IEEE Press*, 207-216.
- [14] Keivanloo, I., Rilling, J., y Zou, Y. (2014). Spotting working code examples. *ACM*, 664-675.
- [15] Irrazábal, E., Garzás, J., & Marcos, E. (2011). Alignment of Open Source Tools with the New ISO 25010 Standard-Focus on Maintainability. In *International Conference on Software and Data Technologies* (Vol. 2, pp. 111-116).
- [16] Mascheroni, M. A., & Irrazábal, E. (2018). Continuous testing and solutions for testing problems in continuous delivery: A systematic literature review. *Computación y Sistemas*, 22(3), 1009-1038.
- [17] Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele University*, 33(2004), 1-26.
- [18] Kitchenham, B., & Pfleeger, S. L. (2002). Principles of survey research. *ACM SIGSOFT Software Engineering Notes*.
- [19] Carruthers, J. A., & Ojeda Rodríguez, C. (2021). Evaluación de conjuntos de datos utilizados

- en la construcción de modelos para la predicción de defectos en clases de proyectos software. In IV Jornadas de Calidad de Software y Agilidad (pp. 7-16).
- [20] Pinto Oppido, J. A., Carruthers, J. A., & Irrazábal, E. A. (2021). Sonar JUploader, aplicación para el análisis, sincronización y actualización automática de proyectos a Sonar Cloud. In IV Jornadas de Calidad de Software y Agilidad (pp. 27-36).

PERSONALIZACIÓN REGIONAL DE BUENAS PRÁCTICAS DE CALIDAD DE SOFTWARE

Emanuel Irrazabal; Gladys Dapozo; Cristina Greiner; Andrea Lezcano Airaldi; Juan Andrés Carruthers; María de los Ángeles Ferraro; María Laura Godoy

Facultad de Ciencias Exactas y Naturales y Agrimensura.
Universidad Nacional del Nordeste
{eirrazabal, alezcano}@exa.unne.edu.ar

RESUMEN

Esta línea se enmarca en el Proyecto F018-2017; una continuación de los proyectos F07-2009, F10-2013 y F018-2017 enfocados en modelos, métodos y herramientas para la calidad del software de Universidad Nacional del Nordeste. Este proyecto aborda los temas emergentes en el área de la calidad de software, en particular, aspectos referidos a la gestión de proyectos en administraciones públicas regionales, la mejora de las pruebas continuas de software en equipos ágiles de desarrollo y el desarrollo de un modelo para la evaluación de visualizaciones con técnicas narrativas o de *storytelling*.

En particular, se está trabajando en el análisis de la transformación digital en la administración pública regional, la validación de un modelo de pruebas continuas de software y la definición de un enfoque para la evaluación de la calidad y mejora asistida de visualizaciones aplicando técnicas narrativas.

Palabras clave: calidad de software, transformación digital, visualización de la información.

CONTEXTO

Las líneas de Investigación y Desarrollo presentada en este trabajo corresponden al proyecto PI-21F05 “Personalización de buenas prácticas de calidad de software en la región”, acreditado por la Secretaría de Ciencia y Técnica de la Universidad Nacional del Nordeste (UNNE) para el

periodo 2022-2025, y a la beca interna doctoral de CONICET otorgada por RESOL-2021-154-APN-DIR#CONICET para el período 2021-2025.

1. INTRODUCCIÓN

Desde el punto de vista de la calidad del software es crítico para las organizaciones entregar productos de calidad de manera aún más rápida. Esto dio lugar al surgimiento de un nuevo enfoque denominado “Entrega Continua de Software”, más conocido en inglés como Continuous Delivery (CD). En este enfoque los equipos mantienen la producción de software en ciclos cortos de tiempo, asegurando que el producto pueda ser lanzado de manera fiable en cualquier momento [1] [2].

Por tanto, es esencial desarrollar un enfoque de priorización de los diferentes aspectos en la calidad del producto software, teniendo en cuenta la forma de trabajo actual de las empresas de desarrollo software. Para adoptar este enfoque, se utiliza el concepto de Tubería de Despliegue (DP - Deployment Pipeline), un estándar para automatizar el proceso de CD [3].

En el grupo de investigación hemos trabajado en el desarrollo de un modelo para la evaluación y mejora de pruebas continuas de software, que se ha utilizado en diez proyectos software con sus correspondientes estudios de caso [4].

Otra de las líneas de trabajo es la personalización de procedimientos de calidad de software en las administraciones

públicas regionales. Se ha venido trabajando con procedimientos de requisitos para empresas yerbateras o el desarrollo de buenas prácticas de toma de requerimientos para sistemas críticos ferroviarios. Esto evidencia la alta velocidad de la adopción de tecnologías digitales en entornos donde el orgullo por lo analógico era y es lo más común [5]. Esta nueva realidad está construida, en parte, por empresas y pequeños equipos de trabajo ágiles [6]. Sin embargo, no es igualmente fértil para todos; las entidades públicas tienen una tradición analógica y de poco uso de las nuevas tecnologías [7]. El aumento de la velocidad de reconversión desde las tareas analógicas y presenciales hacia el desarrollo de servicios digitales integrados debido a la pandemia COVID-19 ha tenido como resultado la transformación digital masiva y desapareja [8].

En el grupo de investigación se está diseñando un conjunto de experimentos para medir la transformación digital en la región, las decisiones que se han tomado y la posibilidad de crecimiento futuro sostenido.

Finalmente, otra línea de investigación es acerca de la narrativa basada en datos, una estrategia efectiva para transmitir información, siendo su propósito estimular la atención y el compromiso (o *engagement*) de la audiencia y facilitar la toma de decisiones [1]. En el contexto actual, donde las posibilidades de toma de datos de distintos sistemas software ha tenido un crecimiento explosivo (Big Data), la visualización de datos proporciona un medio valioso por permitir un análisis asistido (por ejemplo, mediante controles de sumarización, división de los datos, búsquedas contextuales, o gráficos interactivos) [10]. Más aún, dado que muchas veces se utilizan para la toma de decisiones, es importante considerar las perspectivas de los usuarios finales en la construcción de dichas narrativas. Esta interrelación entre: i) las fuentes de datos que dan soporte a una visualización y ii)

los intereses, preferencias y necesidades de los destinatarios de la misma, determina en gran medida la calidad de las narrativas a construir. No considerar estos aspectos suele ocasionar problemas relacionados con: baja calidad del contenido visual, pobre utilización de las fuentes de datos, ineficiencia en el desarrollo del contenido y, en general, poca entrega de valor.

En el contexto actual de desarrollos ágiles, uno de los problemas recurrentes es la poca calidad del contenido visual [11]. Por ello, la implementación de las buenas prácticas de *storytelling* en la construcción de un sistema y su posterior evaluación, pueden tener un impacto significativo en la calidad del producto resultante y en la entrega de valor por parte del sistema a los usuarios finales [12][13].

Se han realizado varios estudios que incluyen las prácticas y pautas generales a seguir para crear visualizaciones efectivas [1][10], [14][15] o [16]. Si bien se han propuesto algunos criterios para evaluaciones parciales, no se encuentran trabajos exhaustivos, herramientas que asistan a ello o propuestas sistemáticas de modelos para evaluar las narrativas basadas en datos.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El objetivo principal del proyecto de investigación presentado es la mejora en la calidad del software y su aplicación en la región del nordeste mediante métodos, herramientas y experimentación para el desarrollo y mantenimiento de software; orientados a aumentar la eficiencia, confiabilidad y seguridad en los distintos ámbitos de aplicación de sistemas informáticos. Los objetivos particulares son los siguientes:

1. Estudiar y profundizar conceptos vinculados con la estimación de software en SCRUM, en particular la experimentación en *story points* y días ideales.

2. Elaborar métodos y herramientas para gestión de proyectos en administraciones públicas regionales.

3. Verificar una metodología de pruebas continuas.

4. Desarrollar un modelo para la medición de calidad de visualizaciones de información teniendo en cuenta buenas prácticas de narrativa digital.

En este marco también se propone el desarrollo de transferencias tecnológicas y de formación de recursos humanos en temas relacionados con la mejora de la calidad de software hacia los equipos de trabajo de la región. Esto facilita los estudios de casos y los cuasi experimentos necesarios para la personalización de los desarrollos. Por otro lado, se busca la especialización de recursos humanos en temas relacionados con la calidad del software a través de la elaboración de tesis de carreras de grado, de posgrado y trabajos de investigación realizados en el marco de becas de investigación para alumnos y/o graduados.

3. RESULTADOS OBTENIDOS/ESPERADOS

Pruebas continuas de desarrollo software

Se ha presentado el modelo de mejora de pruebas continuas en el Simposio Doctoral de la Conferencia Internacional sobre Automatización de Pruebas de Software [17].

Se ha realizado la lectura de la tesis doctoral de un integrante del equipo de investigación en la Universidad Nacional de La Plata [18].

Actualmente se está trabajando con diferentes empresas de desarrollo software para el uso y mejora del modelo.

Transformación digital

Al ser esta una nueva línea de trabajo se está realizando el diseño de un conjunto de

experimentos para medir el grado de transformación digital en las entidades públicas y privadas de la región, haciendo especial hincapié en el grado de implicación de los equipos internos, los proveedores externos, la institucionalización de los cambios y los desafíos a futuro para su mantenimiento. Parte del análisis previo al desarrollo de esta línea se dio a partir de los resultados obtenidos del estudio de caso llevado adelante por el grupo de investigación en [19].

Narrativa digital

En el marco de esta línea se llevó a cabo un estudio de caso [19] con el objetivo de determinar los beneficios de incluir las buenas prácticas de visualización de datos en el desarrollo de un sistema realizado en el contexto actual de crisis sanitaria con tiempos límite restringidos. Para ello, se analizaron 16 gráficos de un sistema de información que gestiona el seguimiento del aislamiento y los permisos de circulación en la cuarentena por la pandemia de COVID-19.

Los resultados indican que el uso de técnicas de narrativa basada en datos contribuye a facilitar el proceso de toma de decisiones mediante la implementación de buenas prácticas de visualización que aumentaron la comprensión y la memorabilidad de los gráficos.

Por otro lado, se realizó un estudio de caso con el objetivo de desarrollar e implementar una aplicación web para un grupo de investigación inserto en una institución académica, utilizando una historia como hilo conductor para capturar y retener la atención de los usuarios [20].

Los resultados evidenciaron en general, una experiencia positiva al navegar el sitio para realizar las tareas propuestas. Esto podría deberse a que se planteó al usuario como el “protagonista” de la historia, apelando de esta manera a la interacción con el sistema para obtener información.

Asimismo, se implementó un modelo basado en el proceso analítico jerárquico en combinación con redes neuronales de propagación hacia atrás para evaluar la calidad de visualizaciones de datos, teniendo en cuenta el conocimiento y experiencia de distintos evaluadores [21]. En base a la literatura estudiada, se elaboró el sistema de índices de evaluación; se seleccionaron diferentes usuarios como expertos evaluadores y se calculó el peso de cada uno. Con los resultados de la evaluación de 20 visualizaciones se entrenó un modelo de red neuronal de propagación hacia atrás. Los resultados de este estudio demostraron que la combinación de métodos refleja la competencia de los expertos y reduce la subjetividad de la evaluación.

Por otro lado, se está realizando una revisión sistemática de la literatura con el objetivo de caracterizar las buenas prácticas en la construcción de visualizaciones de datos y las estrategias de evaluación reportadas al momento. En ese contexto, se finalizó la etapa de ejecución, para la cual se analizaron 96 estudios y se continúa trabajando actualmente en la etapa de reporte.

Dentro de esta línea de trabajo, se espera además, contribuir a la calidad de software a través del diseño y desarrollo de tableros de control con visualizaciones narrativas, bajo la hipótesis de que “la aplicación de estas técnicas mejorará la experiencia del usuario final y contribuirá a una mejor comprensión de la información.”

Como trabajo futuro, se prevé desarrollar un modelo de medición de calidad de visualizaciones a partir de la evaluación de las buenas prácticas implementadas.

4. FORMACIÓN DE RECURSOS HUMANOS

En esta línea de trabajo del Grupo de Investigación sobre Calidad de Software (GICS) están involucrados 4 docentes investigadores, 2 becarios internos doctorales CONICET y 1 becario de

investigación de pregrado.

5. REFERENCIAS

- [1] M. Lehman, "On Understanding Laws, Evolution and Conservation in the Large Program Life Cycle", *J. of Sys. and Software*, vol. 13, pp. 213-221, 1980.
- [2] Randell, Brian. "Fifty Years of Software Engineering-or-The View from Garmisch." arXiv preprint arXiv:1805.02742 (2018).
- [3] Kong, Pingfan, et al. "Automated testing of android apps: A systematic literature review." *IEEE Transactions on Reliability* 99 (2018): 1-22.
- [4] Mascheroni, Maximiliano Agustin, Emanuel Irrazábal, and Gustavo Rossi. "Continuous Testing Improvement Model." 2021 IEEE/ACM International Conference on Automation of Software Test (AST). IEEE, 2021.
- [5] Nadkarni, Swen, and Reinhard Prügl. "Digital transformation: a review, synthesis and opportunities for future research." *Management Review Quarterly* 71.2 (2021): 233-341.
- [6] Thiel, Peter A., and Blake Masters. *Zero to one: Notes on startups, or how to build the future*. Currency, 2014.
- [7] Ismail, Mariam H., Mohamed Khater, and Mohamed Zaki. "Digital business transformation and strategy: What do we know so far." *Cambridge Service Alliance* 10 (2017): 1-35.
- [8] Tekic, Zeljko, and Dmitry Koroteev. "From disruptively digital to proudly analog: A holistic typology of digital transformation strategies." *Business Horizons* 62.6 (2019): 683-693.
- [9] Nussbaumer Knaflic, C. *Storytelling with Data: A data visualization guide for business professionals*. Hoboken, New Jersey: John Wiley & Sons, Ltd., 2015.
- [10] Kosara, R., Mackinlay, J. *Storytelling: The Next Step for Visualization*.

- Computer, Vol 46, Issue 5, 44-50 (2013)
- [11] Nicoletti, M., Schiaffino, S., Díaz Pace, J. A.: An optimization-based tool to support the cost-effective production of software architecture documentation. *J. Softw. Evol. Process.* 27(9): 674-699 (2015).
- [12] Baker, R. *Agile UX Storytelling*. 2017.
- [13] Zimmerman, B. Applying Tufte's principles of information design to creating effective Web sites. *ACM SIGDOC Annu. Int. Conf. Comput. Doc. Proc.*, pp. 309–317, 1997, doi: 10.1145/263367.263406.
- [14] Kosara, R. An Argument Structure for Data Stories. *Proc. Eurographics/IEEE VGTC Symposium on Visualization (EuroVis)*, 2017.
- [15] Tufte, E. R. *The Visual Display of Quantitative Information*. 1983.
- [16] Ram D., et al. Case Report: Rapid Development of Visualization Dashboards to Enhance Situation Awareness of COVID-19 Telehealth Initiatives at a Multi-Hospital Healthcare System. *Journal of the American Medical Informatics Association* (2020).
- [17] Mascheroni, Maximiliano Agustin. Modelo de mejora para pruebas continuas. Diss. Universidad Nacional de La Plata, 2021.
- [18] Mascheroni, Maximiliano Agustin, Emanuel Irrazábal, and Gustavo Rossi. "Continuous Testing Improvement Model." 2021 *IEEE/ACM International Conference on Automation of Software Test (AST)*. IEEE, 2021.
- [19] Lezcano Airaldi A., Diaz-Pace J.A., Irrazábal, E. (2021) Data-driven Storytelling to Support Decision Making in Crisis Settings: A Case Study. *JUCS - Journal of Universal Computer Science* 27(10): 1046-1068. <https://doi.org/10.3897/jucs.66714>
- [20] Lezcano Airaldi, A., Sandoval, D. & Irrazábal, E. (2021). Storytelling Aplicado al Diseño de Sitios Web: Un Estudio de Caso. En *IV Jornadas de Calidad de Software y Agilidad* (pp. 37-46).
- [21] Lezcano Airaldi, A., Acevedo, J., & Godoy, M. L. (2021). Evaluación de la calidad de visualizaciones de datos basada en el Proceso Analítico Jerárquico y Redes Neuronales de Propagación hacia Atrás. En *IV Jornadas de Calidad de Software y Agilidad* (pp. 17-26).

HEURÍSTICAS PARA LA CREACIÓN DE MODELOS DE EVALUACIÓN MULTICRITERIO

Ana Funes, Aristides Dasso

Departamento de Informática
Facultad de Ciencias Físico-Matemáticas y Naturales
Universidad Nacional de San Luis
Ejército de los Andes 950 - 5700 San Luis, Argentina
afunes@unsl.edu.ar, arisdas@unsl.edu.ar

RESUMEN

A partir de la experiencia ganada en la creación de modelos multicriterio LSP en diversos dominios, todos ellos desarrollados en el marco de una línea de investigación en evaluación y comparación de sistemas complejos, hemos podido detectar que una de las tareas que más tiempo consume en todo el proceso de modelado es aquella de agregar preferencias para obtener una preferencia global. Esto se debe, principalmente, a la necesidad de seleccionar operadores adecuados, que representen lo más fielmente posible los criterios del decisor al agregar dichas preferencias. Es así que ha surgido la necesidad de contar con una herramienta computacional que incorpore un cierto grado de conocimiento, ganado a partir de la experiencia, y que ayude al decisor en la tarea de creación de sus modelos, sirviendo de guía en la conceptualización y posterior evaluación.

Para tal fin, en primer lugar, se deberán identificar las principales heurísticas que guían al decisor en la selección de los operadores usados para la creación de las estructuras de agregación de los modelos LSP así como en la determinación de los pesos asignados a las preferencias agregadas, para luego, poder incorporar todo este conocimiento en una aplicación web que integre todo el proceso de modelado y evaluación.

Palabras clave: métodos de evaluación multicriterio, métricas, LSP, AHP.

CONTEXTO

El presente trabajo de investigación se encuentra enmarcado en una línea de investigación del Proyecto de Ciencia y Técnica PROICO 03-2020 “Ingeniería de Software: Estrategias de Desarrollo, Mantenimiento y Migración de Sistemas en la Nube”, de la Universidad Nacional de San Luis. (Director: Daniel Riesco. Acreditado con evaluación externa. Financiamiento: Universidad Nacional de San Luis).

En particular, para esta propuesta de investigación, se busca, además, desarrollarla como una tesis en el ámbito del Departamento de Informática de la Universidad Nacional de San Luis.

1. INTRODUCCIÓN

La Ingeniería de Software, comprende el uso de principios científicos, de ingeniería y de enfoques disciplinados y sistemáticos para la especificación, implementación, operación y mantenimiento de software de alta calidad. Para esto, la Ingeniería de Software adopta un proceso, que abarca una serie de áreas, dentro de las cuales se encuentra la medición de la calidad, no solo del proyecto y del proceso de software sino del producto mismo.

Considerando que la calidad del producto es multidimensional [10][6], existen en la literatura una enorme cantidad de métricas que permiten evaluar cada una de las distintas dimensiones de la calidad de un producto de software [11][7]. Sin embargo, al momento de integrar múltiples atributos en una única medida, por lo general, la mayor parte de las

propuestas apuntan a la aplicación de métodos meramente aditivos, como lo es la ponderación lineal o combinación lineal ponderada (Weighted Linear Combination, WLC), que se basa en la suma ponderada de los atributos.

La WLC tiene un carácter compensatorio, es decir, en el proceso de agregación, el valor bajo de un factor puede ser compensado por otro que ha recibido una puntuación más alta. Además, si bien es un método sencillo, deja de lado la interacción y subordinación entre factores [12]. Por lo tanto, contar con un método cuantitativo, que permita la creación de una función de criterio compleja y que dé solución a estos problemas, resulta una herramienta muy útil a la hora de evaluar o tomar decisiones que involucran múltiples criterios.

El objetivo de la evaluación multicriterio es ayudar en la selección de la “mejor alternativa” que satisfaga los criterios o preferencias del decisor, en un contexto donde existen criterios en competencia y muchas veces en conflicto [18]. En este sentido, LSP [5] ha resultado muy adecuado ya que se trata de un método multicriterio genérico, basado en la Lógica de Preferencia Continua (CPL) [6], para la creación de criterios complejos de evaluación. Se basa en un conjunto de atributos de entrada y permite modelar el carácter compensatorio y/o no compensatorio de los atributos. El mismo ha sido aplicado con éxito en una variedad de dominios [1][2][3][4][9].

En general, y en particular también en el caso de la evaluación de atributos de calidad del software, al realizar una evaluación, se deben considerar las características, atributos y relaciones que sean relevantes para un propósito dado y para una necesidad de información de una categoría de entidad en concreto. Entre las actividades que plantea el método LSP se encuentra el diseño de indicadores, tanto parciales como globales, para obtener el grado de cumplimiento global de las características de alto nivel que se estén evaluando. Básicamente, LSP permite establecer criterios de evaluación,

especificando las propiedades esperadas de un sistema. En este punto, todos los valores de los indicadores elementales pueden agruparse adecuadamente diseñando una Estructura de Agregación (EA) por niveles, que permita obtener una preferencia (valor de indicador global/parcial) de acuerdo a las necesidades y el punto de vista del decisor. El problema principal, en esta actividad, radica en la elección de los operadores adecuados para agregar preferencias, así como en la asignación de pesos a dichas preferencias. No obstante, existe una serie de heurísticas que guían la selección de estos operadores y determinan formas genéricas o patrones de estructuras de agregación, los cuales pretendemos identificar claramente para plasmar en la aplicación web propuesta, que sirva como guía para el decisor.

Por otro lado, un paso casi obligatorio en la evaluación multicriterio es la selección de los pesos que se asignan a cada factor. En el caso de LSP, cada preferencia tiene distintos niveles de importancia y estas diferencias son consideradas al construir la EA por medio de la asignación de pesos a dichas preferencias. A su vez, tal como fundamenta Saaty en [17], “La mente del ser humano no es capaz de considerar todos los factores de un problema y sus efectos al mismo tiempo”, por lo cual, proponemos, además, facilitar la tarea de asignación de pesos, enriqueciendo el método LSP con la integración del método AHP [15][12] en lo que a esta actividad se refiere.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La línea de investigación en la que se enmarca el trabajo presentado, es parte de una investigación sobre la construcción de modelos de evaluación de sistemas complejos y que viene desarrollándose desde hace tiempo en el ámbito del SEG (Software Engineering Group) de la Universidad Nacional de San Luis, donde se han obtenido resultados que han sido plasmados en diversas publicaciones (ver por ejemplo [1][2][3][4][8][9][14]).

Es a partir de la experiencia acumulada en la aplicación del método LSP que ha surgido la necesidad de contar con una herramienta de software que ayude al decisor en la creación de sus modelos. Si bien, al momento, existen diversas aplicaciones que facilitan dicha actividad [8] [13] [14], estas se basan en la creación de la EA desde cero, sin implementar ayudas o guías para la agregación de las preferencias ni la selección de los operadores de agregación. Asimismo, todas ellas dejan librada la asignación de pesos al decisor, realizando un simple chequeo de consistencia en la suma de los pesos asignados. Es por esto que nos hemos propuesto como ejes principales de la presente línea de investigación los siguientes:

- Identificación de patrones genéricos de estructuras de agregación que guíen al decisor en la creación de una EA de preferencias elementales y parciales y en la selección de los operadores CPL involucrados en la misma.

En general, el proceso que sigue el decisor cuando genera la EA consiste en agregar las preferencias elementales en un proceso “bottom-up”, obteniendo una EA que replica mayoritariamente la estructura del AP. Sin embargo, un proceso más sencillo, consiste en categorizar las preferencias en obligatorias, opcionales y deseables para así después agregarlas usando patrones “típicos” para cada caso. En este sentido, un primer soporte que la herramienta debe ofrecerle al decisor es que, una vez que este ya ha creado el Arbol de Preferencias (AP) y ha categorizado las mismas, es brindar el patrón de la EA más adecuada para la situación dada. Por lo tanto, esta primera EA acomodará las preferencias de acuerdo a dicha categorización, conteniendo además operadores CPL típicos en cada nodo de agregación.

Asimismo, considerando que las funciones de agregación son operadores de una lógica continua, el decisor deberá poder refinar los operadores “por defecto” asignados en la estructura genérica, si así

lo desea, de forma coherente con la categorización que se haya realizado de las preferencias. Este proceso semi-automático, pretende reducir de manera considerable el trabajo de creación y puesta a punto de la EA.

- Integración del método AHP en la asignación de los pesos relativos de cada preferencia elemental o parcial de entrada a un operador CPL de agregación.

Para asignar los pesos, se seguirán las 4 etapas del método AHP, pudiendo, en nuestro caso, obviarse la 1^{ra} de las etapas del método (construcción de la estructura jerárquica), dado que la misma se corresponde en forma directa con la creación del AP en el método LSP.

En la 2^{da} etapa, de establecimiento de prioridades, AHP requiere que el decisor evalúe de manera objetiva la importancia de las preferencias. La valoración se realiza mediante la comparación pareada de alternativas, donde se otorgan puntuaciones de 1 a 9, determinando la importancia relativa de una alternativa en relación a otra. En esta etapa se ve la ventaja del método, que evita al decisor tener que asignar todos los pesos de una sola vez.

Luego viene una 3^a etapa de evaluación de consistencia, para verificar que el decisor, al efectuar una gran cantidad de valoraciones de a pares, no haya asignado puntuaciones inconsistentes. Para esta situación, el método AHP cuenta con su índice de consistencia (IC) [16] el cual no debe ser superior al 10%. Finalmente, cuando la asignación devuelve un IC que es válido, se determina el vector de pesos.

- Diseño e implementación de una aplicación web que incorpore los requisitos planteados de los 2 puntos anteriores.

3. RESULTADOS OBTENIDOS/ ESPERADOS

El principal objetivo de este trabajo es la identificación e implementación, en una aplicación web, de patrones genéricos de EA en el método de evaluación multicriterio LSP, así como la integración del método AHP para la asignación de pesos a las preferencias de una AE, todo esto con el fin de facilitar la tarea de elaboración de modelos multicriterio de evaluación cuantitativos, que permitan conceptualizar, medir y evaluar el grado de cumplimiento de un determinado atributo y sub-atributos de manera integral.

4. FORMACIÓN DE RECURSOS HUMANOS

La línea de investigación sobre construcción de modelos de evaluación de sistemas complejos ha producido dos tesis de posgrado en la Maestría en Ingeniería de Software de la Universidad Nacional de San Luis. Una de ellas sobre evaluación de atributos de calidad de sitios de gobierno electrónico [2][3] y la otra sobre el atributo de calidad Accesibilidad en aplicaciones web [8][9], mientras que una tercera tesis de maestría sobre la evaluación del atributo de calidad Aprendizaje se encuentra en desarrollo [1].

La propuesta aquí expuesta, también, tiene como objetivo ser motivo de tesis. En ese sentido, la identificación de los patrones de EA, la integración de AHP y el desarrollo de la aplicación presentada en este trabajo servirán de base para el desarrollo de una tesis en el ámbito del Departamento de Informática de la Universidad Nacional de San Luis.

5. BIBLIOGRAFIA

- [1] J. L. Andrada, A. Funes, A. Dasso. Evaluación del Atributo de Calidad del Software ‘Aprendizaje’ empleando un Método Multicriterio. Anales de SIIO 2021 (JAIIO 2021), Bs. As., Argentina.
- [2] M. Castro. “Análisis de las propiedades y atributos propios de sitios de gobierno electrónico”, Tesis de Maestría en Ingeniería de Software, Departamento de Informática, Universidad Nacional de San Luis, 2010.
- [3] M. Castro, A. Dasso, A. Funes. “Modelo de Evaluación para Sitios de Gobierno Electrónico”, Simposio de Informática en el Estado (SIE) 2009 – 38 JAIIO, Mar del Plata, Argentina, August 26-28, 2009. pp. 200-214.
- [4] A. Dasso y A. Funes, “Threat and Risk Assessment Using Continuous Logic”, Encyclopedia of Organizational Knowledge, Administration, and Technologies, 1st. edition. IGI Global. 2020
- [5] J.J. Dujmović. “Soft Computing Evaluation Logic. The LSP Decision Method and Its Applications”. © 2018 John Wiley & Sons, Inc
- [6] J.J. Dujmović, Preference Logic for System Evaluation. IEEE Transactions on Fuzzy Systems, Vol. 15, No. 6, December 2007.
- [7] Fenton and Pfleeger. Software Metrics: A Rigorous and Practical Approach, Brooks-Cole Publishing, 1998.
- [8] C. Gallardo, A. Funes, H. Ahumada. “Soporte para la Medición y Evaluación de la Accesibilidad al Contenido en Aplicaciones Web”, Anales de ASSE 2019 (JAIIO 2019), Salta, Argentina. pp. 56-70.
- [9] C. Gallardo, A. Funes. “Un Modelo para la Evaluación de la Calidad de la Accesibilidad al Contenido Web”, CONAIISI 2015, Bs. As., Argentina.
- [10] ISO/IEC 25010. <https://iso25000.com/index.php/normas-iso-25000/iso-25010>.
- [11] S. H. Kan. Metrics and Models for Software Quality Engineering, 2nd edition, Addison-Wesley, 2003.
- [12] Malczewski, J. On the Use of Weighted Linear Combination Method in GIS: Common and Best Practice Approaches. Transactions in GIS, 4, 5-22. 2000.
- [13] E. Miranda et al. “NESSy: A new evaluator for software development tool”. 2nd Symposium on Languages, Applications and Technologies. 2013.

- [14] M. Peralta, C. Salgado. “Una Herramienta para la Evaluación de Sistemas”, Tesis de Licenciatura en Ciencias de la Computación, Universidad Nacional de San Luis. 2004.
- [15] T. L. Saaty, *The Analytic Hierarchy Process*. McGraw-Hill, New York. 1980.
- [16] T. L. Saaty, Decision making for leaders, in *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-15, no. 3, pp. 450-452, May-June 1985, doi: 10.1109/TSMC.1985.6313384.
- [17] T. L. Saaty. *What is the Analytic Hierarchy Process?* Springer. 1988.
- [18] H. Voogd. Multicriteria evaluation with mixed qualitative and quantitative data. *Environment and Planning B*, 9, 221-236. 1982.

Interacción Persona-Computadora y Autismo: Interfaces que facilitan la interacción

M. Claudia Albornoz, Mario M. Berón, Germán Montejano

Departamento de Informática/Universidad Nacional de San Luis-U.N.S.L./San Luis/Argentina

Ejército de los Andes 950, Tel: +54 (0266) 4520300; int 2102

{albornoz,mberon,gmonte}@unsl.edu.ar

Resumen

El mundo actual se mueve y se comunica a través de la tecnología, la pandemia del COVID19 introdujo al universo entero a la virtualidad. En este contexto es sumamente importante considerar a las personas con discapacidad.

El avance tecnológico, ha beneficiado a personas con discapacidad. En la actualidad existe una gran variedad de software y herramientas digitales que son de mucha ayuda para ellas/os.

Es fundamental que todas las personas puedan interactuar con los productos tecnológicos: aplicaciones, sistemas de software, sitios de internet. Por lo antes mencionado, lograr que el usuario inexperto pueda comunicarse de manera fluida con los productos tecnológicos es un gran desafío. Como punto de partida, se debe mostrar al usuario una interfaz accesible, amigable, satisfactoria; que le permita realizar trámites y tareas fácilmente.

El presente trabajo describe la línea de investigación que aborda el estudio de cómo lograr interfaces que faciliten la concreción de tareas a personas con discapacidad, en particular aquellas diagnosticadas en el Espectro Autista.

Palabras clave: Interfaz Gráfica de Usuario, autismo, discapacidad, Interacción Persona Computadora.

Contexto

El presente trabajo se enmarca en el Proyecto de Investigación: “Ingeniería de

Software: Estrategias de Desarrollo, Mantenimiento y Migración de Sistemas en la Nube”. Dirección: Daniel Riesco código P-032020 de la Facultad de Ciencias Físico, Matemáticas y Naturales, de la Universidad Nacional de San Luis. Dicho proyecto es la continuación de diferentes proyectos de investigación a través de los cuales se han logrado importantes vínculos con diferentes universidades a nivel internacional y nacional. Además, se encuentra reconocido por el Programa de Incentivos.

Introducción

En los últimos años, con la irrupción a nivel mundial del Covid19, la gran mayoría de las personas se ha visto en la necesidad de interactuar con la tecnología. Los productos tecnológicos forman parte de la vida cotidiana de las personas; tanto en la actividad profesional, laboral; como también en el hogar o para concretar tareas como extraer dinero de un cajero, comunicarse con otras personas (videollamadas, mensajes, redes sociales, etc.), solicitar un turno en instituciones gubernamentales o sistemas de salud pública, educación en modalidad virtual, etc. Para el usuario inexperto, sin conocimientos, estas tareas pueden ser complicadas y engorrosas si las interfaces con las que interactúan no facilitan la tarea.

Se puede definir *interfaz* como: conjunto de formas y métodos que posibilitan la interacción de un sistema con los usuarios utilizando formas gráficas e imágenes. Con formas gráficas se refiere a botones, íconos,

ventanas, fuentes, etc. los cuales representan funciones, acciones e información [1].

Si bien se habla de usuario inexperto, también es necesario tener muy en cuenta a las personas adultos mayores y personas que presentan alguna discapacidad, ya sea motora, visual, auditiva, verbal, intelectual.

Son numerosas las herramientas y aplicaciones diseñadas para personas con discapacidad, por ejemplo, herramientas que convierten audio en texto para personas hipoacúsicas, otras que convierten texto en audio para personas ciegas, también existen aquellas que facilitan la comunicación y educación de personas con diagnóstico TEA (Trastorno del Espectro Autista) [2,3]. Para que las personas con discapacidad puedan beneficiarse con el uso de estas herramientas, las mismas deben contar con interfaces que faciliten la interacción.

Sin dudas, el diseño de estas herramientas cuenta con interfaces adaptadas a sus necesidades, pero el objetivo es facilitar la interacción de las personas con discapacidad con cualquier producto de software a través de interfaces accesibles.

Desde hace tiempo se trabaja arduamente para lograr la inclusión de personas con discapacidad en diferentes ámbitos. La industria del software no es la excepción. Este trabajo tiene como tema central la discapacidad, dado lo extenso de la temática se hará referencia a la discapacidad intelectual, particularmente el Trastorno del Espectro Autista (TEA).

I.P.C. en personas con TEA

La Interacción Persona Computadora (IPC) es el punto de encuentro de las ciencias humanas, como la psicología, la pedagogía; con las ciencias exactas, como la informática. La IPC es la ciencia que analiza los dos aspectos fundamentales: el humano y la computadora en conjunto [4].

A modo de definición, se puede decir que: la interacción persona-computadora (IPC) es

el intercambio observable de información, datos y acciones entre un humano y la computadora, y viceversa [13].

El Trastorno de Espectro Autista (TEA) se caracteriza por una amplia gama de trastornos del desarrollo, entre los que se encuentran la capacidad imaginativa, el comportamiento, la comunicación y la interacción social recíproca [14]. Una de las dificultades más notorias del TEA, es en el ámbito de la socialización. Se les dificulta entender el mundo de los pensamientos, la interacción social, emociones, creencias, deseos, tanto de los demás como de ellos mismos. Es posible, que el niño desde muy temprana edad presente dificultades, al estar en contacto con los demás, para comprender miradas, para emplear y entender expresiones faciales, gestos, diferentes tonos de voz, entre otros.

Aún no se tiene dimensión de la enorme ayuda que puede brindar la IPC (Interacción Persona-Computadora) a todas las personas diagnosticadas con TEA; dado que el entorno visual es fundamental para ellos, el buen diseño de la interfaz gráfica de usuario les puede abrir caminos y facilitar la interacción social. Esto les puede otorgar una gran autonomía en su diario vivir.

Es fundamental que el usuario pueda interactuar con la interfaz, de eso depende el éxito del producto. La interfaz es el medio de comunicación entre la persona y el producto; es el medio que se utiliza para la interacción. Si la interfaz entorpece o no facilita la interacción, provoca que el usuario cometa errores, no concrete las tareas, se frustre y el producto se considere un fracaso.

El entorno visual, gráfico es muy importante dado que las personas con TEA presentan dificultad para comprender la información presentada como texto. Si bien ellos aprenden a leer sin problema, se les presenta dificultad al procesar, comprender lo que leen [5]. Las personas con TEA asimilan la información asociándola con

imágenes, he aquí la importancia del componente gráfico.

Es de público conocimiento el gran beneficio que se obtiene al aplicar las TIC (Tecnologías de la Información y las Comunicaciones) en el aprendizaje, la enseñanza, la educación; especialmente personas diagnosticadas con TEA.

Si bien el alumnado general muestra interés por trabajar con las TIC, los alumnos con TEA se sienten particularmente atraídos. Las TIC se adecúan a las características de los alumnos y facilitan su aprendizaje; además permiten adaptar los contenidos curriculares dentro de una misma tarea para todo un grupo de alumnos [6].

Es inaceptable que las TIC sean ajenas a los procesos educativos; se las considera un sustento de la sociedad actual por lo tanto no pueden quedar fuera del aula. Las TIC deben ser usadas en el aprendizaje y la enseñanza, facilitando, transformando los procesos educativos; permitiendo el acceso del amplio espectro de las sociedades a una educación de calidad [7]. Hay quienes creen que las TIC condenan a las personas con autismo a encerrarse aún más en su propio mundo, a profundizar más su autismo; pero se ha comprobado que esto no es real, muy por el contrario, les ayuda a combatir el aislamiento, favorecen la interacción social [12], les ayudan a compartir actividades educativas, de comunicación, de diversión, a trabajar en equipo. Dentro de este contexto, se fortalecen habilidades sociales y comunicativas, que a su vez posibilitan las relaciones interpersonales y reconocimiento de emociones [11].

Hasta hace algunos años atrás se consideraba que las personas diagnosticadas con TEA debían recibir una educación especial, diferencial; sin tener la oportunidad de ser incluidos en la escuela normal. Hoy en día es completamente diferente, las escuelas en la actualidad ejercen una educación inclusiva la cual se ve altamente beneficiada

con la intervención de las TIC y de la computadora en el aula. Así, todas las personas con discapacidad pueden ser incluidas en la educación regular.

Entre las grandes cualidades que presentan las TIC se pueden mencionar:

- ✓ favorecen el trabajo autónomo.
- ✓ se constituyen como refuerzo positivo en la enseñanza y el aprendizaje de los alumnos con TEA.
- ✓ permiten cierto grado de error y así se disminuye la frustración, permiten repetir tareas y acciones favoritas de los alumnos, el alumno aprende disfrutando [6].

La interacción con la computadora no sólo ayuda en la educación y el aprendizaje de personas con TEA; también los puede ayudar en la vida social. Si se utiliza la computadora como medio de comunicación adecuado y funcional, tienden a disminuir sus conductas sociales inadecuadas [9]. Además, el trabajo de la computadora puede estimular a la persona con autismo a un uso activo del lenguaje [10].

Es importante mencionar que, para que el uso de las TIC sea efectivo, es necesario que presenten una interfaz motivadora que integre la información en todos los formatos (imágenes, íconos, texto, audio, etc.), que sea clara en los contenidos, que incorpore colores, movimiento, elementos visuales. En particular se espera que presente un diseño divertido. Inútil será un software con fines educativos que incorpore las últimas novedades técnicas, que incluya mucha tarea su diseño y elaboración, pero que presente una interfaz aburrida o poco atractiva para los alumnos con TEA.

El diseño de la interfaz es sumamente importante ya que es la parte visible del producto, a través de la cual se interactúa. El diseño y construcción de la interfaz inicia con la identificación de los requerimientos del usuario. Luego se deben crear y analizar las tareas. Durante la fase de diseño es importante seguir un conjunto de principios,

que permitan identificar los objetos y acciones que deberán estar visibles, así se constituye la base del prototipo de la interfaz de usuario. A medida que el diseño avanza, se pueden hacer las modificaciones y ajustes necesarios; por último, es necesario evaluar la calidad del resultado [8].

Líneas de Investigación, Desarrollo e Innovación

Este artículo es parte de un extenso y minucioso trabajo de investigación respecto a los factores que influyen el diseño y construcción de interfaces gráficas de usuario. La investigación ha mostrado que es fundamental reconocer los distintos usuarios finales: diferenciando su grupo etario (niños, adolescentes, adultos, adultos mayores), personas con y sin experiencia, personas con y sin discapacidad.

Hasta el momento, en esta línea de investigación, se están llevando a cabo las siguientes tareas:

- Revisión Sistemática de la literatura referente al tema de diseño y construcción de Interfaces Gráficas de Usuario; con el fin de evidenciar la importancia de su diseño y cómo lograr una óptima y satisfactoria interacción con los usuarios. Haciendo una mención especial respecto al rol del usuario en el diseño de la Interfaz.
- Estudio de Métodos de Evaluación Multicriterios. Esta tarea tuvo como objetivo seleccionar un método de evaluación para evaluar interfaces la calidad de las interfaces gráficas de usuario. Hasta el momento se ha encontrado que el método LSP (Logic Scoring of Preference) se adapta muy bien para la evaluación de las GUI.
- Elaboración de Criterios de Evaluación. Esta tarea se lleva a cabo con el fin de establecer cuáles son las características que las interfaces gráficas de usuario deben tener para que sean consideradas

adecuadas para su utilización por parte del usuario. Estos criterios serán la entrada al método de evaluación multicriterio mencionado en el ítem precedente.

- Estudio de Reglas, Técnicas y Principios de Diseño de GUI. Esta tarea tiene como finalidad proveer la información necesaria para que los equipos de desarrollo de GUI puedan desarrollar GUIs adecuadas para el usuario.
- Análisis de la Influencia Emocional, el objetivo es determinar cómo influye el estado emocional de los usuarios en la interacción con las computadoras y, a su vez, cómo el diseño de la Interfaz influye en el estado emocional del usuario.

Resultados y Objetivos

El diseño y construcción de la interfaz gráfica de usuario no debe considerarse como una tarea secundaria, sin importancia. Asiduamente se cree que la tarea del equipo desarrollador sólo se reduce a cumplir con los requerimientos del usuario, sin considerar características fundamentales que debe poseer la interfaz; parte fundamental del producto, dado que es la parte visible y con quien interactúa y se comunica el usuario.

En este contexto se puede decir que se está desarrollando un Trabajo Final de Especialización donde se aborda la temática de esta línea de investigación en profundidad.

La investigación se centra en identificar los factores que influyen el diseño y construcción de interfaces gráficas de usuario. Para esto se ha hecho un estudio sistemático de la literatura; esto ha permitido identificar que uno de los principales factores es la Interacción Persona-Computadora; en particular, personas con discapacidad, haciendo foco en personas dentro del espectro autista (TEA).

Como resultado, esta investigación pretende determinar cómo lograr una GUI 'usable', comprensible, satisfactoria; con el

análisis de los diferentes tópicos que se involucran en la Interacción Persona-Computadora.

Futuros trabajos: i) Enriquecer el estudio sistemático de la literatura; ii) Implementar un método de evaluación multicriterio para evaluar GUIs, iii) Definir criterios de evaluación de interfaces gráficas de usuario, iv) Elaborar un manual con las Reglas, Técnicas y Principios de Diseño de GUI y v) Profundizar el estudio sobre la influencia emocional.

Formación de Recursos Humanos

Los progresos obtenidos en esta línea de investigación sirven como base para el desarrollo de tesis de posgrado, ya sea de doctorado, maestrías o especializaciones en Ingeniería de Software; y desarrollo de trabajos finales de las carreras Licenciatura en Ciencias de la Computación, Ingeniería en Informática e Ingeniería en Computación de la Universidad Nacional de San Luis, en el marco de los Proyectos de Investigación.

Referencias

- [1] <http://www.alegsa.com.ar/Dic/gui.php>
- [2] Bonilla, M., & Chaskel, R. (2016). Trastorno del espectro autista. Programa de Educación continua en Pediatría. Sociedad Colombiana de Pediatría, 15(1), 19-29.
- [3] Reynoso, C., Rangel, M. J., & Melgar, V. (2017). El trastorno del espectro autista: aspectos etiológicos, diagnósticos y terapéuticos. Revista médica del instituto mexicano del seguro social, 55(2), 214-222.
- [4] Moralejo, L., Sanz, C., and Pesado, P. (2014). Paradigmas de interacción hombre-máquina, un análisis enfocado al ámbito de la educación especial. Revista Ruedes, Año 3- No 5- ISSN: 1853-5658, pages 85-101.
- [5] González Navarro, A., Freire Prudencio, S., Gil, D., Martos Pérez, J., Jordanova, V., Cerga Pashoja, A., Shishkova, A., and Evans, R. (2014). First: una herramienta para facilitar la comprensión lectora en el trastorno del espectro autista de alto funcionamiento. Rev Neurol, pages 129-135.
- [6] Martínez, J. L., Pagán, F. J. B., García, S. A., and Máiquez, M. C. C. (2016). Las tecnologías de la información y comunicación (tic) en el proceso de enseñanza y aprendizaje del alumnado con trastorno del espectro autista (tea). Revista Fuentes, (14):193-208.
- [7] Peñuela, J. R. J. (2011). Posibilidades educativas de las tic para la población infantil autista. TELOS 87: Nuevos perfiles profesionales para la comunicación digital, 87:9-11.
- [8] Carreño-León, M, Leyva-Carrillo, A, Carreño, M, & Sandoval, A. Consideraciones para el diseño de interfaces de usuario en aplicaciones para niños con autismo. AVANCES SOBRE REFLEXIONES, APLICACIONES Y TECNOLOGÍAS INCLUSIVAS, 70.
- [9] De la Vega, M. E. and Koon, R. A. (2000). La computadora en la intervención de niños y adolescentes con autismo.
- [10] Guzmán, G., Putrino, N., Martínez, F., & Quiroz, N. (2017). Nuevas tecnologías: Puentes de comunicación en el trastorno del espectro autista. Terapia psicológica, 35(3), 247-258.
- [11] Muñoz, R., Nöel, R., Kreisel, S., & Mancilla, F. (2012). Proyect@ Emociones: software para estimular el desarrollo de la empatía en niños y niñas con trastornos del espectro autista. Nuevas ideas en informática educativa, TISE, 59-64.
- [12] Pedro López, A. D. (2020). Diseño de la interacción en aplicaciones educativas para personas con Trastorno del Espectro del Autismo (Master's thesis).
- [13] Díaz, F. J., Harari, I., and Amadeo, A. P. (2013). Guía de recomendaciones para el diseño de software centrado en el usuario. Editorial de la Universidad Nacional de La Plata (EDULP).
- [14] Carreño-León, M., Leyva-Carrillo, A., Carreño, M., & Sandoval, A. (2019). Consideraciones para el diseño de interfaces de usuario en aplicaciones para niños con autismo. In *Avances sobre reflexiones, aplicaciones y tecnologías inclusivas* (pp. 70-76). Conaic.

HACIA LA ADAPTACIÓN DE SCRUM PARA INCORPORAR CALIDAD DE DATOS EN ÁMBITO DEL DESARROLLO ÁGIL

Carrizo Claudio*, Javier Saldarini*, Gonzalo Castillo*, Sofía Bovo*, Rocío Cortese*, Angélica Caro#, Carlos Salgado+, Alberto Sanchez+, Mario Peralta+

*Facultad Regional San Francisco – Universidad Tecnológica Nacional

Av. de la Universidad 501 - San Francisco - Córdoba - Tel. 03564-421147

{cjcarrizo77, saldarinijavier, gonzalocastilloar, sofibovo501, rociocortese24}@gmail.com

#Departamento de Ciencias de la Computación y Tecnologías de la Información - Facultad de Ciencias Empresariales, Universidad del Bío-Bío

Casilla 447, 3780000, Chillán, Chile

{mcaro}@ubiobio.cl

+Departamento de Informática - Facultad de Ciencias Físico-Matemáticas y

Naturales Universidad Nacional de San Luis

Ejército de los Andes 950 – C.P. 5700 – San Luis – Argentina

{csalgado, alfanego, mperalta}@unsl.edu.ar

RESUMEN

En la actualidad, la mayoría de las organizaciones utilizan Sistemas de Información (SI) para llevar adelante sus procesos de negocio y toma de decisiones. Los sistemas informáticos son parte de los SI, y uno de sus elementos principales son los datos. Estos comúnmente son utilizados en diferentes ámbitos, como por ejemplo, procesos de negocio, gestión del conocimiento, toma de decisiones, explotación de la información, analítica de datos, etc. Por este motivo, resulta lógico pensar que los datos deben tener un nivel de calidad adecuado, de manera que las organizaciones puedan cumplir sus objetivos estratégicos. En el ámbito del desarrollo ágil, no es común definir, especificar e implementar requisitos de calidad en relación a los datos, y la metodología ágil Scrum no es la excepción. El objetivo de esta línea de investigación consiste en realizar una propuesta de adaptación del Framework de Scrum, para que se considere la definición, especificación e implementación de requisitos de calidad de datos, usando

como referencia la Norma ISO/IEC 25000. La principal contribución de esta investigación consiste en ayudar a garantizar la calidad de los datos en contexto del desarrollo ágil, particularmente cuando se usa Scrum.

Palabras clave: Requisitos de Calidad de Datos – ISO 25000 - Desarrollo Ágil – SCRUM -

CONTEXTO

El presente trabajo se enmarca dentro del siguiente proyecto de Investigación:

- Ingeniería de Software: Estrategias de Desarrollo, Mantenimiento y Migración de Sistemas en la Nube – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto N° P-03-2020. Dicho proyecto es la continuación de diferentes proyectos de investigación a través de los cuales se ha logrado un importante vínculo con distintas universidades a nivel nacional e internacional. Además, se encuentra reconocido por el programa de Incentivos.

1. INTRODUCCIÓN

Los Sistemas de Información (SI) forman parte de un sistema mayor denominado “Sistema de la Empresa” (ISO/IEC 25030:2007), por lo tanto, la mayoría de las organizaciones hace uso de dichos SI para llevar adelante sus procesos de negocio y la toma de decisiones. Según (ISO/IEC 25030:2007), en un modelo de sistema, los sistemas informáticos son parte de un SI. En (M. Piattini et. al, 2011) se menciona no solo la relevancia que tienen los SI para las organizaciones, sino que también se hace hincapié acerca de la importancia que tiene la calidad de los sistemas informáticos, para que las organizaciones puedan cumplir sus objetivos estratégicos. En (ISO/IEC 25030:2007) también se menciona que los sistemas informáticos están compuestos por los siguientes 4 elementos: hardware, sistemas operativos, software de aplicación y datos. Este último, es un elemento de gran importancia, ya que es fundamental para la toma de decisiones objetivas en todos los niveles de una organización (Javed, B. & Hussain, S., 2003; Levy S., 2004; Naveh, E. & Halevy, A., 2000). Las organizaciones modernas consideran a los datos como el activo más importante, y como uno de sus recursos estratégicos (Olson, J., 2002; Tayi, G. & Ballou, D., 1998), debido a que a partir de los mismos, se pueden llevar adelante proyectos relacionados a la generación de conocimiento (Ponjuán, Gloria, 2006), proyectos de integración (Crm, 2002), dar soporte a la toma de decisiones y explotación de la información (Inmon, W.H., 1996), afrontar proyectos de analítica de datos (V. M. Schönberger & K. Cukier, 2013), etc.

Según lo expuesto anteriormente, es fundamental que los datos que gestiona un SI tengan un nivel de calidad adecuado,

esto significa, que estén libres de errores y que posean características relevantes, para que el dato sea de utilidad para quienes lo consuman. Estas características permiten establecer un lenguaje común con el propósito de focalizar problemas de calidad de datos y oportunidades de mejora (Javed, B. & Hussain, S., 2003; Naveh, E. & Halevy, A., 2000).

En (ISO/IEC 25012: 2008) se define la calidad de datos como el “grado en que las características de los datos satisfacen necesidades implícitas y establecidas cuando son usados en condiciones específicas”.

En la literatura existen varias propuestas que abordan la calidad de los datos desde la perspectiva de características (Richard Y. Wang, D.M.S., 1996; ISO/IEC 25012: 2008), una de ellas corresponde a una reconocida Familia de Normas denominada ISO/IEC 25000, también conocida como SQuaRE (Systems and software engineering-Systems and software Quality Requirements and Evaluation), la cual propone un marco para la definición de requerimientos de la calidad de software/datos y evaluación de la calidad del software/datos, apoyados por un proceso de medición de la calidad de software/datos. En lo que respecta a la calidad de datos, SQuaRE incluye la Norma ISO/IEC 25012, la cual proporciona un modelo general de características de calidad de datos, que están contextualizadas desde 2 puntos de vista: inherente y dependiente del sistema. Por otro lado, también incluye la Norma ISO/IEC 25024, la cual brinda un conjunto de medidas asociadas a las características del modelo de calidad de datos.

En cuanto al abordaje de la calidad de datos en el ámbito del desarrollo de software, en (R. Wang, 1998) se destaca la

importancia de incorporar aspectos de calidad de datos en fases tempranas del desarrollo, específicamente en fase de desarrollo de requisitos. Si bien existen trabajos que abordan la incorporación de aspectos de calidad de datos en etapas tempranas del proceso de desarrollo de software (A. Caro et. al, 2013; A. Rodríguez y A. Caro, 2013; C. Guerra-García et. al, 2016; Y. Fernanda Torres Alonso, 2020), se han encontrado escasas propuestas en el ámbito del desarrollo de Software, a través del uso de Metodologías Ágiles. En la industria del software, existen varias metodologías de este tipo, entre ellas Scrum (Ken Schwaber & Jeff Sutherland, 2020), Extreme Programming y Kanban, siendo Scrum la metodología ágil más adoptada y utilizada a nivel mundial. En el contexto de Scrum, también existen escasas propuestas que aborden la incorporación de aspectos de calidad asociados a datos, desde la perspectiva de la especificación temprana e implementación de requisitos de calidad de datos, en un escenario de ambiente controlado de desarrollo.

Lo descripto anteriormente, nos motivó a iniciar una línea de investigación cuyo objetivo principal es proponer una adaptación del Framework de Scrum, a fin de incorporar aspectos de calidad de datos, desde la perspectiva de la especificación e implementación de requisitos de calidad de datos. Estos requisitos se obtienen a partir de un proceso de análisis y definición, por medio del uso de las Normas ISO/IEC 25030, ISO/IEC 25012, e ISO/IEC 25024.

La adaptación propuesta incluirá 3 aspectos principales:

1. La creación de un nuevo rol dentro de Scrum, el cual tendrá la tarea de gestionar la calidad de los datos.

2. La definición de un método o técnica que permita guiar el trabajo de especificación y verificación de cumplimiento de implementación de requisitos de calidad de datos.
3. El desarrollo de herramientas que faciliten la aplicación del método o técnica propuesta.

El aporte principal de esta línea de investigación consiste en potenciar el marco de trabajo de Scrum, de manera que se pueda abordar la calidad de datos desde el momento en que se inicia el desarrollo de un producto de software.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Los principales ejes de trabajo de esta línea de investigación se detallan a continuación:

- Realizar una revisión sistemática de la literatura acerca de la temática calidad de datos en contexto del desarrollo de software en general, y en el ámbito del uso de la metodología ágil Scrum en particular.
- Realizar un estudio de la Familia de Normas ISO/IEC 25000, en particular de las Normas ISO/IEC 25030, ISO/IEC 25012, e ISO/IEC 25024.
- Realizar un estudio de la última versión del manifiesto ágil de Scrum, para comprender como funciona la metodología ágil.
- Definir un rol dentro de Scrum, el cual será responsable de velar por la calidad de los datos, durante el desarrollo del producto de software.
- Definir métodos o técnicas que permitan guiar la especificación y

verificación de implementación de requisitos de calidad de datos, basados en SQuaRE.

- Realizar pruebas de concepto para realizar los ajustes que sean necesarios.
- Desarrollar herramientas que permitan facilitar el uso y/o aplicación del método o la técnica definida.

3. RESULTADOS OBTENIDOS/ESPERADOS

Resultados Obtenidos

- Informe que contiene el resultado de la revisión sistemática de la literatura.
- Informe que contiene el resultado del estudio realizado sobre la Familia de Normas ISO/IEC 25000.
- Conjunto de requisitos de calidad de datos, obtenidos a partir de un proceso de análisis y definición, a través de ISO/IEC 25012 e ISO/IEC 25024.
- Informe del estudio realizado al manifiesto ágil de Scrum.

Resultados Esperados

- Método o técnica definida y ajustada, luego de la prueba de concepto realizada.
- Herramientas de uso libre que permitan automatizar el método o la técnica propuesta.
- Nueva versión del Framework de Scrum, que integre el método o técnica propuesta, y las herramientas.

FORMACIÓN DE RECURSOS HUMANOS

Esta línea de investigación se trabaja desde el Grupo de Investigación

denominado "Calidad de Software y Datos", el cual pertenece a la UTN Facultad Regional San Francisco, Córdoba. El equipo de trabajo está constituido por: 1 Director de Proyecto, 1 Co-Director, 4 docentes investigadores, 1 graduado y 1 estudiante, todos de la especialidad Sistemas de Información. Cabe destacar que también se están llevando a cabo algunas tesis de grado y trabajos finales, en el marco de las Ingenierías de las distintas universidades.

5. BIBLIOGRAFÍA

A. Caro, A. Fuentes, A. Soto. "Desarrollando sistemas de información centrados en la calidad de datos", Revista Chilena de Ingeniería, vol.21, pp. 54-69, 2013.

C. Guerra-García, I. Caballero, M. Cardenas-Juarez, R. Samano-Robles, "Adding quality in the user requirements specification: A first approach", 2016 Eleventh International Conference on Digital Information Management (ICDIM). IEEE.

CrM. Como Mejorar Las Relaciones Con Los Clientes. Pearson Educación, 2002.

Extreme Programming. Disponible en <https://www.agilealliance.org/glossary/xp>.

Inmon, W.H. "The data warehouse and data mining." Communications of the ACM, vol. 39, no. 11, Nov. 1996, pp.

ISO/IEC 25000 Systems and software engineering- Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models.

ISO/IEC 25030:2007, Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Quality requirements., ISO, 2007.

ISO/IEC 25012: 2008. Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE) -- Data quality model.

- ISO/IEC 25024:2015. Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE). Measurement of data quality.
- Javed, B. & Hussain, S. (2003). Data quality – A problem and an approach. Wipro Technologies.
- Kanban. Redacción APD. Disponible en <https://www.apd.es/metodologia-kanban>.
- Ken Schwaber & Jeff Sutherland, 2020. La Guía Definitiva de Scrum: Las Reglas del Juego <https://scrumguides.org/docs/scrumguide/v2020/2020-Scrum-Guide-Spanish-European.pdf>.
- Levy, S. (2004). Model Documents and forms for Organizing and Maintaining a Data Quality Program.
- M. Piattini, F. O. G. I. García y F. Pino, Calidad de sistemas de información, 9 ed., España: RaMa, 2011.
- Naveh, E. & Halevy, A. (2000). A hierarchical framework for a quality information system. Total Quality Management, Vol.11, No. 1, p 87-111.
- Olson, J. (2002). Data Profiling: The Data Quality Analyst's Best Tool. DM Direct, December. DMReview.com.
- Ponjuán, Gloria. (2006). Gestión de Información en las Organizaciones. Editorial Félix Varela.
- R. Wang, "A Product Perspective on Total data Quality Management". Communications of the ACM. Vol. 41, Issue 2, pp. 58-65. 1998.
- Richard Y. Wang, D.M.S., Beyond accuracy: what data quality means to data consumers. Journal of Management Information Systems, 1996. 12(4): p. 5-33.
- Tayi, G. & Ballou, D. (1998).Examining Data Quality. Communications of the ACM. Vol. 41, No. 2.
- Viktor Mayer-Schönberger, Kenneth Cukier. Big data: la revolución de los datos masivos. Turner, 2013 - 320 páginas.
- Y. Fernanda Torres Alonso, "Especificación de requerimientos de software con un enfoque de calidad de datos", Univesridad Autónoma de San Luis, Potosí, 2020.

Uso de Técnicas de Design Thinking para mejorar los Sistemas de producción en PyMEs

Panessi Walter¹, Challiol Cecilia², Ortiz Claudia¹

¹ Departamento de Ciencias Básicas, Universidad Nacional de Luján

[wpanessi; cortiz]@unlu.edu.ar

² LIFIA, Facultad de Informática, UNLP y CONICET

ceciliac@lifia.info.unlp.edu.ar

Resumen

En este trabajo se presenta una sub línea de investigación que se desprende de un proyecto presentado en 2020 con duración de cinco años denominado “*Interfaces no convencionales aplicadas a la captura de datos en procesos productivos industriales de las PyMEs*”. En la línea de investigación que se presenta en este trabajo participan docentes y estudiantes de la *Universidad Nacional de Luján* y docentes de la *Universidad Nacional de La Plata*. Esta línea de trabajo tiene como objetivo proponer un enfoque para mejorar la interacción humano-computadora en sistemas de producción de PyMEs de modo de hacer más efectiva la integración del software con la cultura organizacional de este tipo de empresas. Para ello se pretende utilizar las lecciones aprendidas desde la Interacción Humano - Computadora utilizando además técnicas de Design Thinking.

Palabras clave:

Design Thinking; Interacción Humano-Computadora; PyMEs.

Contexto

Este trabajo es una línea de investigación enmarcada en un proyecto de investigación y desarrollo denominado “*Interfaces no*

convencionales aplicadas a la captura de datos en procesos productivos industriales de las PyMEs” radicado en el *Departamento de Ciencias Básicas de la Universidad Nacional de Luján*, que tiene como duración 6 años¹ (2020-2025). Para esta línea, ya se cuenta con un tema de tesis aprobado para la *Maestría en Ingeniería de Software de la Universidad Nacional de La Plata* con fecha de finalización en el corriente año.

Introducción

Las PyMEs son empresas que, por su rango, desarrollan un menor volumen de actividad que las grandes empresas. Según Farinelli (Farinelli, 2007) esto les aporta una mayor flexibilidad para adaptarse a los cambios del mercado y emprender proyectos innovadores lo que redundaría en una buena fuente generadora de empleo y constituyen una parte sustancial de la economía (Gatto and Yoguel, 1993). Apoyar su tecnificación significa entonces un aporte importante para toda la comunidad.

En Argentina, las PyMEs son predominantemente empresas familiares. Una alta proporción de ellas son firmas concebidas y administradas por grupos familiares, tanto en cuanto a la propiedad de la empresa como por el tipo de gestión

¹ Cinco años con uno de prórroga producto de la pandemia.

empresarial. Esto repercute en diversos aspectos del funcionamiento económico y productivo de la firma como, por ejemplo, el nivel de centralización en la toma de decisiones y en la importancia asignada a la experiencia personal en temas vinculados con la gestión comercial y productiva, se tejen lazos entre los empleados y la alta dirección muy fuertes, etc. (Farinelli, 2007).

Los estudios realizados por los métodos de la Organización del Trabajo, en los cuales se recolectan datos respecto de *qué se hace, dónde, cómo y con qué*, permiten aportar trazabilidad a los productos; el mayor problema que emerge es determinar cómo se capturan esos datos, sobre todo en el marco de las PyMEs. Existen muchos métodos, algunos manuales, otros computarizados (Kume, 1989). Los métodos manuales requieren de un proceso de digitalización posterior para que sean realmente útiles (Patrascanu, 2016). Los métodos computarizados se basan en capturas por teclados de computadoras, uso de scanner de códigos de barra y/o pantallas táctiles (Patrascanu, 2016), (Munim et al., 2020). Todo esto requiere de una interacción entre los usuarios (empleados a los cuales se le asigna la tarea de registro) y la computadora; y de disciplina para no olvidar registrar datos y ser precisos en los registros. Si los sistemas trabajan con información incorrecta o imprecisa, los cálculos en los que se basan el resto de las operaciones no serán correctos. Es de interés mencionar que los métodos computarizados son el foco de estudio de esta línea.

El estudio de la Interacción Humano-Computadora (IHC, Human-Computer Interaction - HCI) es la rama dentro de las ciencias de la computación que estudia las capacidades y limitaciones de interacción

del hombre, las computadoras y de la relación entre ambos (Card et al., 1980). En (Munim et al., 2020) se describe que *“... la integración de HCI en los sectores comercial y empresarial puede desempeñar un papel fundamental en la eficiencia, el rendimiento y la productividad de la automatización. En este sentido, el análisis de usabilidad de los sistemas de gestión de la producción industrial se convierte en una cuestión importante para desarrollar un sistema de automatización industrial...”*. Por otro lado, en (Martig et al., 2004) se establece que *“La incorporación de nuevas tecnologías en los sistemas de control y los avances en las tecnologías de recolección y comunicación de datos han impactado en la forma en que los operadores interactúan con los sistemas”*. Además, ambos trabajos (Martig et al., 2004), (Munim et al., 2020) coinciden que en la actualidad hay una gran cantidad de software disponible que provee una manera cómoda y flexible de visualizar procesos industriales; sin embargo, la mayoría de estas herramientas se basan en una representación mímica del proceso sin aprovechar los avances realizados en las áreas de Visualización de Información y en Interacción Humano-Computadora.

Construir software usable para PyMEs representa un reto adicional, ya que implica conocer la cultura organizacional de forma personalizada (dado que cada empresa familiar tiene distintas dinámicas de funcionamiento) y además empatizar individualmente con los stakeholders (quienes pueden tener experticias muy variadas dentro de cada PyME); de manera de generar productos que sean usables y maximicen el éxito de la captura de datos. Starostka (Starostka, 2014) realiza un estudio para analizar el impacto de Design

Thinking² en la cultura organizacional de empresas de diseño en Suecia y Polonia. Si bien el estudio realizado fue exploratorio y los estudios se realizaron en grandes empresas, han concluido que propiciar la cooperación entre los diseñadores de los productos y los Design Thinkers³ parece un buen punto de partida para obtener cambios organizacionales generales. Sugieren que su trabajo puede ser un buen punto de partida para proyectos de investigación más profundos sobre el papel de Design Thinking en la construcción de la cultura organizacional innovadora de diferentes tipos de organizaciones.

Acorde a lo antes descrito, surge la motivación de utilizar técnicas de Design Thinking para empatizar con los involucrados en los procesos de producción de PyMEs industriales en Argentina; para poder recolectar los datos adecuados con el fin de mejorar tanto los artefactos software existentes en estas empresas, de modo que sean más usables, como así también identificar la posibilidad de generar nuevos.

En base a la motivación emerge el objetivo de esta línea que consiste en proponer un enfoque para mejorar la interacción humano-computadora en sistemas de producción de PyMEs de modo de hacer más efectiva la integración del software con la cultura organizacional de estas empresas con el fin de hacer que las formas

en que se obtienen los datos de los hechos productivos se adapten a las personas para minimizar errores y, por consiguiente, hacer más exactos sus sistemas de control de producción y, al mismo tiempo, hacer más confortable la forma de trabajar de las personas.

Hipótesis de Trabajo

Se partirá de la hipótesis de que es posible mejorar la interacción humano - computadora en sistemas de producción de PyMEs de modo de hacer más efectiva la integración del software con la cultura organizacional de este tipo de empresas mediante un enfoque de Design Thinking, el cual permita obtener como resultado posibles aspectos de mejoras como así también propuestas concretas para llevarlos a cabo.

Para comprobar la hipótesis se comenzará por un proceso de medición de la productividad y se conducirán estudios tendientes a medir el grado de satisfacción del personal en su trabajo. Estos estudios comparados con los mismos datos, pero obtenidos luego de aplicar el enfoque propuesto por esta línea de investigación permitirán demostrar si es posible la mejora esperada usando este tipo de abordaje.

2 Design Thinking. Brown (Brown, 2008) define: “Design Thinking es una disciplina que usa la sensibilidad y métodos de los diseñadores para hacer coincidir las necesidades de las personas con lo que es tecnológicamente factible y con lo que una estrategia viable de negocios puede convertir en valor para el cliente, así como en una gran oportunidad para el mercado”. Cross (Cross, 2011) brinda una definición más general planteando que “el término Design Thinking se

refiere a los procesos cognitivos del trabajo de diseño, o las habilidades de pensamiento y las prácticas que los diseñadores utilizan para crear nuevos artefactos o ideas y resolver problemas”.

3 Design Thinker. Brown (Brown, 2008) menciona este término para referirse a las personas que participan en la puesta en práctica de *Design Thinking*, las cuales deberían tener las siguientes habilidades la empatía, el pensamiento integrador, la experimentación y la colaboración.

Objetivos

El objetivo de esta línea de investigación es proponer un enfoque para mejorar la interacción humano-computadora en sistemas de producción de PyMEs de modo de hacer más efectiva la integración del software con la cultura organizacional de este tipo de empresas. Este enfoque tiene como fin encontrar las estrategias más adecuadas para capturar los datos que surgen de los hechos productivos y, por consiguiente, hacer más exactos sus sistemas de control de producción. La captura de estos acontecimientos productivos no es una tarea trivial, e implica lograr una cercanía con los involucrados en cada PyME; es decir, se vuelve fundamental comprender y entender, no solo las necesidades de los sistemas sino también las capacidades y forma de trabajo de los involucrados. Acorde a esto, se utilizarán técnicas de Design Thinking como parte del enfoque para lograr empatizar con la cultura organizacional de cada PyME.

Líneas de Investigación, Desarrollo e Innovación

Dado el contexto interdisciplinario de este trabajo, se abordarán varios temas de investigación que se mencionan a continuación.

- *Interacción Humano-Computadora*: Se estudiarán los factores humanos y tecnológicos que intervienen en la interacción. Las distintas formas de interacción, sus ventajas y desventajas. Los factores que hacen a la tecnología usable para los humanos sin generar rechazos.
- *Design Thinking*: Se analizarán las distintas visiones de este término como así también los distintos frameworks conceptuales que

sirven de guías para poner en práctica este tipo de diseño; destacando ventajas y desventajas. Recursos existentes que pueden ser usados en la etapa de empatizar con las personas, recursos para la etapa de ideación como así también de prototipado para probar los resultados que se van diseñando; analizar cuáles de estos se adecuan mejor a PyMEs.

- *Abordajes en empresas*: Se estudiarán antecedentes de enfoques de Design Thinking relacionadas con Usabilidad, Diseño Centrado en el usuario y otras formas de mejorar y evaluar la interacción humano-computadora. Antecedentes del uso de Design Thinking para mejorar procesos industriales. Experiencias de investigaciones y desarrollo en relación a la Interacción Humano-Computadora en Industrias.

Los temas mencionados anteriormente serán relevantes para dar contexto y poder incorporar las bases conceptuales no sólo para proponer el enfoque de esta línea de la investigación, sino también para poner en práctica los casos de estudios.

Es de interés destacar que el trabajo experimental será realizado una vez que el enfoque que surja en este trabajo este definido; el cual permitirá mediante la utilización de distintos recursos/técnicas de Design Thinking lograr empatizar con los stakeholders de los sistemas de producción en PyMEs, para poder recolectar los datos adecuados para mejorar tanto software existente en estas empresas como así también identificar la posibilidad de generar nuevos artefactos de software.

El trabajo experimental consiste en la puesta en práctica del enfoque propuesto, poniendo especial foco en la etapa de empatizar con las personas mediante el uso

de recursos/técnicas de Design Thinking, sobre los stakeholders relacionados con el sistema de producción. Esto permitirá evaluar el enfoque y poder obtener una retroalimentación para enriquecer el mismo; y además poder identificar posibles recomendaciones en cuanto a la puesta en práctica de este tipo de enfoque en PyMEs.

Se realizará la puesta en práctica del enfoque con las siguientes dos PyMEs:

- PyME - Caso de estudio 1: Empresa metalúrgica productora de puertas estampadas de acero inyectadas con poliuretano expandido. Se trata de una empresa en la que trabajan 60 personas aproximadamente, dirigida por sus dueños. Un emprendimiento familiar que ha crecido hasta convertirse en una empresa altamente tecnificada. Se inicia en la industria metalúrgica en el año 1987. Posee una planta de 14500 metros cuadrados.
- PyME - Caso de estudio 2: Planta Piloto productora de alimentos de la Universidad Nacional de Luján. En ella se desarrollan la mayor parte de las actividades que realiza el Centro de Investigación, Docencia y Extensión en Tecnología de Alimentos (CIDETA). En esta Planta Piloto se encuentran instaladas las siguientes varias líneas de proceso como el procesamiento de productos lácteos, de dulces y conservas frutihortícolas, de productos panificados.

Las PyMEs, por sus características propias, poseen aspectos únicos y por lo tanto es factible encontrar diferencias significativas dentro de cada una respecto

a su cultura organizacional. Por consiguiente, es posible que los recursos para empatizar (y obtener datos significativos) sean diferentes en cada caso. Esto lleva a la necesidad de empatizar seleccionando los recursos de Design Thinking que mejor se adecúen a cada situación. Se espera poder explorar una gama de recursos diferentes y observar la relación entre los recursos y los datos que se obtienen a partir de los mismos.

Una vez que se evalúen los resultados del uso de los diferentes recursos de la etapa de empatización, se podrá determinar cómo seguir con las siguientes etapas del enfoque (por ejemplo: idear, prototipar) y qué recursos podrían ajustarse mejor para cada caso.

Al finalizar la puesta en práctica del enfoque con estos dos casos de estudio, se espera haber probado un conjunto de recursos de Design Thinking que permita establecer recomendaciones respecto a su utilidad para conducir abordajes de este tipo en el ámbito de las PyMEs, y de esta manera poder aportar en la conducción de proyectos de software de producción exitosos en términos de usabilidad en PyMEs.

Formación de Recursos Humanos

Se planifica la realización de una tesis de maestría del alumno de Posgrado de la Universidad Nacional de La Plata, Walter F. Panessi. La propuesta de tema de tesis fue aprobada en 2021 y lleva por título: *“Enfoque para mejorar la Interacción Humano-Computadora en sistemas de producción en PyMEs mediante técnicas de Design Thinking”*. Adicionalmente, se planifica también la realización de un trabajo de tesina de grado, con tema aprobado y que lleva por título

“Aplicaciones de pensamiento de diseño como motor para encontrar soluciones creativas al seguimiento de la producción en PyMes” del estudiante de la carrera Licenciatura en Sistemas de Información de la Universidad Nacional de Luján, Jose Emanuel Rodriguez. Se evalúa además la concreción de otros trabajos de licenciatura y la incorporación de otros profesores, docentes auxiliares y estudiantes en actividades de investigación.

Referencias

- Brown, T. (2008). Design thinking. Harvard Business Rev. Harvard Business Review, 86, 84-95.
- Card, S. K.; Moran T. P.; Newell, A. (1980). The keystroke-level model for user performance time with interactive systems. Communications of the ACM. 23 (7): 396–410.
- Cross, N. (2011). Design thinking: Understanding how designers think and work,” Berg, Oxford. New York.
- Farinelli, S. (2007). Sistemas Locales de Innovación: 'Las empresas Pymes metalmeccánicas de Tandil (1995-2005). Universidad Nacional del Centro de la provincia de Buenos Aires, 2007
- Gatto, F. Yoguel, G. (1993). El Desafío de la Competitividad. La Industria Argentina en transformación. Cáp. IV CEPAL, Alianza Editorial. 1993
- Kume, H. (1989) Métodos estadísticos para el mejoramiento de la calidad. Asociación Argentina de Ex Becarios de la ABK y AOTS.
- Martig, S.; Castro, S; Larrea, M (2004), Diseño de Interfaces Industriales, Proceedings of the IV Congreso Argentino de Enseñanza de la Ingeniería: 331-339, Instituto Tecnológico de Buenos Aires.
- Munim, K.M.; Islam, I.; Rahman, M.M.; Nazrul Islam, M. (2020) Adopting HCI and Usability for Developing Industry 4.0 Applications: A case study. In Proceedings of the 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh.
- Patrascanu, L. (2016). Estudio de los beneficios económicos a raíz de la implantación de captura inalámbrica de datos en planta con la ayuda del ERP Expertis en una empresa manufacturera del sector del composite (REFISA, S.L.). Universitat de Lleida. Facultat de Dret i Economia. 2016
- Starostka, J. (2014). Design and design thinking in building an innovative organizational culture. Journal of Intercultural Management, 6(4-1).

Creación de Software para personas con Discapacidad usando Design Thinking

Ortiz Claudia¹, Challiol Cecilia², Panessi Walter¹

¹ Departamento de Ciencias Básicas, Universidad Nacional de Luján
[cortiz; wpanessi]@unlu.edu.ar

² LIFIA, Facultad de Informática, UNLP y CONICET
ceciliac@lifia.info.unlp.edu.ar

Resumen

Este trabajo surge como una de las sub líneas de investigación de un proyecto aprobado en 2020 denominado “*Interfaces no convencionales aplicadas a la captura de datos en procesos productivos industriales de las PyMEs*”. El objetivo de esta línea es proponer un abordaje de Diseño Centrado en las Personas (DCP) para la creación de software que sea más amigable para personas con discapacidad. En particular, el abordaje de DCP que se decide utilizar para alcanzar el objetivo mencionado es Design Thinking (DT); ya que DT hace principal hincapié en la etapa de empatización para comprender las necesidades reales de las personas.

Palabras clave:

Diseño Centrado en las Personas, Design Thinking, Discapacidad.

Contexto

La línea de investigación presentada en este trabajo se encuentra enmarcada en el proyecto de investigación y desarrollo denominado “*Interfaces no convencionales aplicadas a la captura de datos en procesos productivos industriales de las PyMEs*”, radicado en el Departamento de Ciencias Básicas de la Universidad Nacional de Luján, que tiene como duración 6 años¹ (2020-2025). Para esta línea de investigación se cuenta con un tema de tesis aprobado para la Maestría en Ingeniería de Software de la Universidad Nacional de La

Plata, por Resolución HCD 610/21. Esta tesis se denomina “*Diseño centrado en las personas: el desafío de empatizar*” y tiene fecha de finalización en el corriente año.

Introducción

El Diseño Centrado en el Usuario (DCU) como se enuncia en [Hassan Montero and Ortega Santamaría, 2009] “*es un enfoque de diseño cuyo proceso está dirigido por información sobre las personas que van a hacer uso del producto*”. Por lo tanto, un abordaje de DCU para el desarrollo de software se focaliza en el relevamiento de información de los usuarios para determinar requerimientos funcionales y no funcionales que debe poseer el software; en la mayoría de los casos para esto se utilizan distintas técnicas de recolección de información [Rubin and Chisnell, 2008].

Es de interés mencionar que la gran mayoría de las técnicas de recolección de información no logran reflejar lo que vivencian realmente algunas personas en su vida diaria, como ser aquellas que tienen una condición de discapacidad (por ejemplo, visión/movilidad reducida, problemas en la motricidad o el habla). Esto genera que los diseñadores/ desarrolladores de software no lleguen a empatizar con las necesidades reales de las personas con discapacidad, generando que las soluciones obtenidas no sean realmente usables o las personas se frustren al usarlas [Kahraman, 2020]. Por ejemplo, en [Molina-López and Medina Medina, 2021] se menciona que en los

¹ Cinco años con uno de prórroga producto de la pandemia.

videojuegos el grupo más desfavorecido son las personas con discapacidad visual (que incluye a las personas con ceguera del color), ya que en este tipo de aplicaciones tanto los colores como la velocidad de reacción juegan un rol preponderante que es fundamental para lograr una buena experiencia de uso; sin embargo en el diseño de la mayoría los videojuegos no se consideran las necesidades reales de las personas con discapacidad, y por lo tanto ellas se llevan una gran frustración al querer usarlos.

En [Oliveira et al., 2016] se plantea que la accesibilidad es considerada como un problema de implementación, por lo que se tiene en cuenta cuando el producto está totalmente desarrollado, lo que genera trabajos adicionales y costos extras para hacerlo accesible. Esta es la visión que tienen la mayoría de los desarrollos de software actuales; por lo tanto, se considera importante involucrar al usuario en estadios tempranos de diseño, para no “sufrir” las consecuencias de la incorporación de los mismos en instancias finales. Acorde a esto, se vuelve fundamental dar comienzo al diseño con una etapa de empatización. Como se menciona en [Da Silva and Goncalves Ferrerira, 2020] considerar la empatía puede contribuir a lograr un mundo más accesible.

La importancia de empatizar con las personas con discapacidad se da en todos los ámbitos de la vida cotidiana; por ejemplo, en aspectos relacionados con la arquitectura [Hutton and Maguire, 2021] o en el campo de la salud [Sosa et al., 2018], [Altman et al., 2018].

En [Bennett and Rosner, 2019] se analizan actividades de empatía destinadas a comprender a las personas con discapacidades, el autor pone en tela de juicio la simulación práctica de las “deficiencias” y reclama interacciones más directas. Más aún, en [Waller et al., 2015] sugieren que muchos diseñadores nunca se reúnen con sus usuarios y a menudo solo utilizan personajes ficticios (a veces usando el recurso Personas) para entender a su población de

usuarios. No hay que perder de vista que en algunos casos la interacción directa no es una tarea trivial; en relación a esto, en [Capellen and Andersson, 2021] se fomenta para el co-diseño el uso de artefactos con capas digitales personalizables para facilitar el proceso de interacción con personas con una comunicación verbal reducida.

Hace más de diez años que se ha detectado la necesidad de involucrar activamente a las personas con discapacidad en las investigaciones sobre ellas [Mankoff et al. 2010] reflexión que se continua analizando en la actualidad, poniendo de manifiesto que no se ha encontrado aún la solución a esta necesidad [Spiel et al., 2020]. Esta visión de involucrar a las personas como centro del proceso de diseño se denomina Diseño Centrado en las Personas (DCP) [Norman, 2013], el cual emerge como una evolución del DCU. En el DCP se debería considerar a las personas en todas las etapas de diseño teniendo como premisa garantizar el éxito del producto que se está creando. Sin embargo, como se mencionó anteriormente, en relación a las personas con discapacidad esto sigue siendo una meta pendiente [Spiel et al., 2020].

Acorde a lo antes descrito se busca una solución de abordaje de DCP que ponga el foco en la empatización, y además involucre a las personas con discapacidad en todas las etapas de diseño. La disciplina de Design Thinking (DT) [IDEO, 2015] aborda el DCP, haciendo foco en la importancia de empatizar adecuadamente para que las soluciones que se generen reflejen las necesidades reales de las personas. Existen distintas visiones del término DT, por ejemplo, en [Kimbell, 2011] se menciona que este término puede referirse a: un proceso cognitivo, a una teoría general de diseño, o un recurso de la organización. En ese sentido Cross [Cross, 2011] menciona que “*El término Design Thinking se refiere a los procesos cognitivos del trabajo de diseño, o las habilidades de pensamiento y las prácticas que los diseñadores utilizan para crear*

nuevos artefactos o ideas y resolver problemas”, lo que nos lleva a pensar que puede ser utilizado para dar solución a la generación de productos y/o servicios en diferentes ámbitos; uno de estos es la creación de software. Cabe mencionar que DT está muy poco explorado en soluciones de software orientadas a personas con discapacidad; y esto también es parte de la motivación de la línea de investigación presentada en este trabajo. Como se menciona en [Corso and Challiol, 2020] existen diferentes frameworks conceptuales que pueden usarse de guía para abordar DT, cada uno de estos frameworks divide el proceso de diseño en distintas etapas iterativas; por ejemplo, algunos definen tres fases mientras que otros definen cinco. Más allá de la cantidad de fases en las que cada framework divide el proceso de diseño todos coinciden a grandes rasgos en que se debe realizar una fase de empatización, luego otra de ideación, para finalmente llegar al prototipado/testeo. Para cada fase pueden usarse diferentes recursos/técnicas o una combinación de estos, en [Hehn et al., 2018] se identifican 172 posibles recursos que se pueden utilizar. Sin embargo, como se destaca en [De Paula et al., 2019] usar los recursos más apropiados para cada dominio es uno de los factores de éxito del proceso de diseño. Actualmente, como se menciona en [Corso and Challiol, 2020] no existe una guía clara que permita determinar cuál es la mejor forma para seleccionar los recursos más adecuados para conducir DT, más aún en los dominios menos explorados como el diseño de software para personas con discapacidad. Este tema abierto forma parte de la motivación de esta línea de investigación.

Objetivos

El objetivo de la línea de investigación es proponer un abordaje de DCP para la creación de software para personas con discapacidad, donde el foco central es empatizar adecuadamente para

identificar sus necesidades reales, y luego continuar el proceso de diseño involucrándolos en todas las etapas. Para esto se explorarán distintos recursos/técnicas de DT para determinar cuáles son más adecuados para cada situación o cómo estos se pueden combinar para recolectar información más precisa. Con este abordaje se espera identificar tanto requisitos funcionales como no funcionales para el software a desarrollar para estas personas.

Actualmente, las personas con discapacidad en la mayoría de los diseños/desarrollos de software participan recién en las etapas más avanzadas como, por ejemplo, pruebas de funcionalidad o evaluación de usabilidad. El abordaje propuesto en esta línea de investigación busca poner a las personas con discapacidad como actores centrales desde las etapas tempranas de diseño.

Líneas de Investigación, Desarrollo e Innovación

Las líneas de investigación que se abordarán son:

- *Software para personas con discapacidad.* En particular, software para personas con síndrome de down, desafíos motrices/intelectuales y daltonismo. Se hará principal hincapié en la usabilidad de los sistemas actuales orientados a dar soporte a las discapacidades mencionadas, y cómo estas personas se ven involucradas en el desarrollo de estos softwares. Analizar las dificultades propias de cada discapacidad en relación a los sistemas de software actuales.
- *Frameworks conceptuales y recursos/técnicas de DT.* Relevar y analizar distintos frameworks conceptuales que se usan para DT, para poder determinar qué aspectos podrían ser considerados. Relevar y analizar recursos/técnicas de DT, por ejemplo, mapa de empatía, lluvia de ideas, diagrama de prioridades, selección por clasificación, entre otras. Determinar cuáles son más apropiados para empatizar, y cuáles podrían utilizarse para otras etapas de diseño como idear y/o prototipar.

- *Empatizar con personas con discapacidad.*
Relevar el uso de distintas técnicas de empatización con personas con discapacidad. Técnicas que se utilizan en otras disciplinas como, por ejemplo, arquitectura o salud.

El trabajo experimental permitirá la puesta en práctica del abordaje propuesto, logrando resultados que permitan obtener una retroalimentación para enriquecer el mismo; y además poder brindar recomendaciones en cuanto a la puesta en práctica de variados recursos para las discapacidades en las que se hará foco en los casos estudio.

Se realizará la puesta en práctica del abordaje con los siguientes tres casos de estudio:

- ❖ Caso de estudio 1: *foco en la condición del síndrome de down.*
- ❖ Caso de estudio 2: *foco en los desafíos motrices/intelectuales.*
- ❖ Caso de estudio 3: *foco en la condición de daltonismo.*

Dado que cada persona es única, es factible encontrar diferencias significativas dentro de cada grupo; ya sean de comunicación, necesidades, capacidades, deseos, al momento de desenvolverse ante una aplicación y/o expresar lo que podría acontecer al interactuar con algún recurso; esto lleva a la necesidad de empatizar seleccionando los recursos de DT que mejor se adecúen a cada situación. Pero para seleccionar estos recursos se tiene que conocer a la persona, para poder luego proceder con la selección de los mismos; es decir, se deberán tener varios encuentros con cada persona que participará en los casos de estudio. Dependiendo de las características de los grupos experimentales, se podrían poner en práctica simulaciones que nos permitan acercarnos a sus vivencias como se implementa en [Sosa et. al, 2018].

Puede acontecer que el vínculo anterior con las personas involucradas en cada caso de estudio sea distinto; requiriendo en algunos casos un encuentro inicial para que ambas partes se

conozcan, y empezar desde ahí un vínculo que permita poner en práctica el abordaje. Se espera que esta reunión permita apreciar cómo la persona habla, se comunica y expresa; para poder luego acompañar mejor cuando haya que abordar cada recurso. Al establecer un vínculo de confianza (esto puede llevar uno o más encuentros) se dará inicio a la instrumentación de entrevistas informales focalizadas en datos que sirvan para determinar recursos con los cuales proseguir con la etapa de empatización con más profundidad. Se podría continuar conociendo más de cada persona usando recursos como: cuestionarios donde se formularán preguntas sencillas relacionadas con su experiencia de uso de algún producto, sus sensaciones, sus gustos y deseos futuros. En aquellos casos en los que se dificulte la comunicación, se pondrán en práctica técnicas de observación de uso, con actividades dirigidas. La selección de estos recursos será muy dinámica y se irá definiendo de forma personalizada para cada caso de estudio.

Una vez que se analicen los resultados del uso de los diferentes recursos de la etapa de empatización, se podrá determinar cómo seguir con las siguientes etapas del abordaje (idear, prototipar) y qué recursos podrían ajustarse mejor para cada persona.

Finalizada la puesta en práctica del abordaje, se espera haber probado un conjunto de recursos de DT que permitan establecer recomendaciones respecto a su utilidad para conducir un abordaje de DCP con base en la “empatía” con personas con discapacidad.

Finalmente, es importante destacar el contexto pandémico actual, que tendrá que considerarse al momento de llevar a cabo los casos de estudio, para saber si los encuentros podrán llevarse a cabo en forma presencial o habrá que pensar en otra forma de vinculación; lo cual puede llevar a ajustar algunos recursos, por ejemplo, a la modalidad virtual.

Formación de Recursos Humanos

En relación a la formación de recursos humanos, uno de los miembros del proyecto está actualmente realizando una tesis para la *Maestría en Ingeniería de Software de la Universidad Nacional de La Plata*. Se espera que esta línea de investigación permita poder incorporar docentes, auxiliares y estudiantes en actividades relacionadas con la temática. Se proyecta contar con la realización de, al menos, una tesina de grado de la *carrera Licenciatura en Sistemas de Información de la Universidad Nacional de Luján*.

Referencias

- Altman, M.; Huang, T.; Breland, JY. (2018). Design Thinking in Health Care. *Prev Chronic Dis* 15, E 117.
- Bennett, C.; Rosner, D. (2019). The Promise of Empathy: Design, Disability, and Knowing the "Other". In *Proceedings of the 2019 Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, USA, Paper 298, pp. 1–13.
- Cappelen, B.; Andersson, AP. (2021). Trans-Create - Co-Design with Persons with Severe Disabilities. *Stud Health Technol Inform* 282, pp. 87-101.
- Corso, C.L.; Challiol, C. (2020). Meta-framework para diseñar experiencias de Design Thinking. In *Proceedings of IEEE ARGENCON 2020*, IEEE, Chaco, pp. 1-8.
- Cross, N. (2011). *Design thinking: Understanding how designers think and work*, Berg, Oxford. New York.
- Da Silva, J.; Goncalves Ferreira, M. (2020). For an inclusive design. *Braz. J. of Develop, Curitiba* 6(7). pp. 44878-44888.
- De Paula, D.; Dobrigkeit, F.; Cormican, K. (2019). Doing it Right-Critical Success Factors for Design Thinking Implementation, In *Proceedings of the International Conference on Engineering Design*. England, vol. 1, pp. 3851-3860.
- Hassan Montero, Y.; Ortega Santamaría, S. (2009). Informe APEI sobre Usabilidad. Gijón: Asociación Profesional de Especialistas en Información.
- Hehn, J.; Uebernickel, F.; M. Herterich (2018). Design Thinking Methods for Service Innovation-A Delphi Study, Pacific Asia Conference on Information Systems (PACIS). Japan, Art. 126.
- Hutton, K.; Maguire, M. (2021). How Empathetic modelling positively influences Architects empathy, informing their Inclusive Design-Thinking. *Ergonomics & Human Factors* 2021, Eds R Charles & D Golightly, CIEHF.
- IDEO.org. (2015). *The Field Guide to Human-Centered Design* (1st ed.). Recuperado de <https://www.designkit.org/resources/1> (último acceso: 12-10-2021).
- Kahraman, E. (2020). *Inclusive Design Thinking: Exploring the obstacles and opportunities for individuals and companies to incorporate inclusive design* (Dissertation).
- Kimbell, L. (2011). Rethinking design thinking: Part I, *Design and Culture* 3, pp. 285-306.
- Mankoff, J.; Hayes, G.; Kasnitz, D. (2010). Disability studies as a source of critical inquiry for the field of assistive technology. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility (ASSETS '10)*. Association for Computing Machinery, New York, USA, pp. 3–10.
- Molina Lopez, J.; Medina Medina, N. (2021). Un enfoque para el diseño inclusivo de videojuegos centrado en jugadores daltónicos. *Revista de la Asociación Interacción Persona Ordenador (AIPO)* 2(1), 25-37.
- Oliveira, R.; Silva, L.; Leite, J.C.S.P.; Moreira, A. (2016). Eliciting accessibility requirements an approach based on the NFR framework. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, pp. 1276-1281.
- Rubin, J.; Chisnell, D. (2008). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. Wiley Publishing, Inc.
- Sosa, L.B.; Molina, M.N.; Sáenz, L. (2018). Métodos y técnicas de investigación aplicados en proyectos de diseño para la salud. *Seminario de Investigación en Diseño* 9, pp. 258-267.
- Spiel, K.; Gerling, K.; Bennett, C.; Brulé, E.; Williams, R.; Rode, J.; Mankoff, J. (2020). Nothing About Us Without Us: Investigating the Role of Critical Disability Studies in HCI. In *Proceedings of Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, USA, pp. 1–8.
- Waller, S.; Bradley, M.; Hosking, I.; Clarkson, P. (2015). Making the case for inclusive design. *Applied Ergonomics* 46, pp.297-303.

ATRIBUTOS DE CALIDAD APLICADOS A LA METODOLOGÍA MEDUC_AR PARA EL DESARROLLO DE APLICACIONES CON RA EN EDUCACIÓN

Nélida Raquel Cáceres, Ana Carolina Tolaba, María del Pilar Gálvez, Natalia María del Huerto Flores

Facultad de Ingeniería - Universidad Nacional de Jujuy
Ítalo Palanca 20 San Salvador de Jujuy – 0388 4221576
nrcaceres@fi.unju.edu.ar

RESUMEN

El propósito del proyecto presentado en este trabajo, es incorporar a la metodología Meduc_AR aspectos de calidad según la norma ISO/IEC 25010, entre otras. Meduc_AR es una metodología para el desarrollo de aplicaciones de Realidad Aumentada (RA) destinadas al ámbito educativo. La metodología fue diseñada por el Grupo de Investigación y Desarrollo en Ingeniería de Software (GIDIS) de la Facultad de Ingeniería de la UNJu y consiste en tres fases: Análisis del Problema, Elección de la Solución y Evaluación de la Aplicación. Entre algunas de sus características se destaca el desarrollo iterativo de las fases que la comprenden y el trabajo en equipo entre los desarrolladores y docentes.

A partir de la identificación de los aspectos de calidad se procederá al refinamiento de la metodología Meduc_AR de modo de poder continuar enriqueciendo las características destacadas de esta metodología.

Palabras claves: Metodología de desarrollo, Software de RA, Realidad Aumentada, Ingeniería de Software, Calidad de software.

CONTEXTO

Este trabajo de investigación está comprendido en el proyecto “*Estudio*

comparativo e identificación de vacancias en metodologías para el desarrollo de aplicaciones con RA en educación (Meduc_AR)”. El mismo fue aprobado por la Secretaría de Ciencia y Técnica de la Universidad Nacional de Jujuy como proyecto categoría B (código D/B041) y se encuentra bajo incentivo.

1. INTRODUCCIÓN

Azuma [1] define Realidad Aumentada (RA) como una tecnología que combina contenido real y virtual, es interactivo en tiempo real y se registra en 3D. La RA propone superponer o añadir, en tiempo real, elementos virtuales al campo de visión de una persona, de modo tal que la información requerida esté presente de una manera natural para el usuario.

El funcionamiento de las aplicaciones de RA, tienen tres subsistemas fundamentales [2]: visualización (salida), ubicación de objetos virtuales en el mundo real (registro) y métodos de interacción (entrada).

El ámbito educativo ha adoptado la RA de manera permanente entre sus recursos tecnológicos. Existen diferentes propuestas de aplicaciones educativas que utilizan RA, las mismas para su creación emplean diferentes metodologías de desarrollo.

Una metodología está formada por fases, cada una de las cuales se puede dividir en sub-fases, que guiarán a los desarrolladores de

sistemas a elegir las técnicas y herramientas más apropiadas en cada momento del proyecto de desarrollo y también a planificarlo, gestionarlo, controlarlo y evaluarlo [3].

Una técnica, es un conjunto de procedimientos y recursos que forman parte del proceso de desarrollo de software.

Una herramienta es un conjunto de instrumentos que se utilizan para desempeñar un oficio o un trabajo determinado. En el desarrollo de software, una herramienta es el instrumento mediante el cual pueden llevarse a cabo técnicas determinadas por una metodología.

Para realizar un análisis de las diferentes metodologías en el desarrollo de aplicaciones con RA en la educación se realizó una comparación de diferentes trabajos [4], [5], [6] y [7]. A partir de esto se observó el empleo de metodologías ágiles y/o combinación de metodologías de desarrollo de software. Es decir, que no existe una metodología diseñada específicamente para el desarrollo de aplicaciones con RA destinada a la educación.

Es por ello que GIDIS en [8] propuso una metodología para el desarrollo de aplicaciones educativas usando RA. Dicha propuesta es iterativa y fomenta el trabajo en equipo entre los desarrolladores y el plantel educativo. La metodología propuesta posee 3 fases (figura 1):

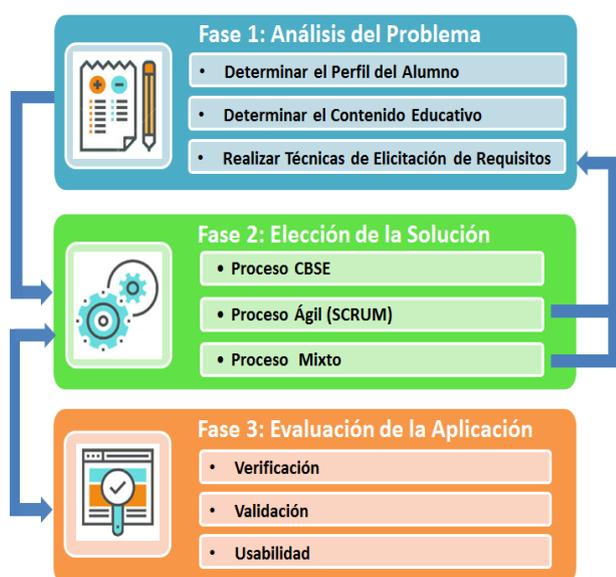


Figura 1. Metodología Meduc_AR [8]

- *Fase 1 - Análisis del Problema:* en donde se identifican los usuarios destinatarios y receptores de la aplicación con RA.
- *Fase 2 - Elección de la Solución:* se determina la viabilidad de la aplicación solicitada y la forma de realizarla, fundamentándose en los requerimientos del usuario.
- *Fase 3 - Evaluación de la Aplicación:* se realiza la verificación y validación de la aplicación desarrollada.

En esta última fase es importante destacar que el objetivo es la obtención de una aplicación con calidad. En [9] se expresa que la calidad del producto software se puede interpretar como el grado en que dicho producto satisface los requisitos de sus usuarios aportando de esta manera un valor. Los requisitos de funcionalidad, rendimiento, seguridad, mantenibilidad, entre otros, se encuentran representados en el modelo de calidad, el cual está definido por la ISO/IEC 25010 y para el cual se definen características de calidad como la adecuación funcional, eficiencia de desempeño, compatibilidad, usabilidad, fiabilidad, seguridad, mantenibilidad y portabilidad.

Considerando que según ISO/IEC 9241-11 la usabilidad de la aplicación, es el nivel con el que un producto se adapta a las necesidades del usuario y puede ser utilizado por el mismo para lograr metas con efectividad, eficiencia y satisfacción en un contexto específico de uso y que todo esto es lo que se busca mediante la metodología Meduc_AR se considera la incorporación a la metodología la evaluación de la usabilidad de la aplicación desarrollada.

En el trabajo de Zumaquero [10] precisamente se realiza la evaluación de usabilidad para RA, por lo que se debe rescatar del mismo el método utilizado para este trabajo, similar al sistema tradicional, en donde se afirma que si se cumplen las expectativas del usuario y las sensaciones del mismo son buenas, su grado de satisfacción será alto y el sistema puede considerarse válido, siempre y cuando éste sea también funcionalmente correcto. En el concepto especificado para usabilidad se detalla lo siguiente:

- *Efectividad*. La precisión con la que el usuario puede alcanzar los objetivos especificados.
- *Eficiencia*. Cantidad de recursos utilizados en relación con la exactitud con la cual los usuarios alcanzan objetivos.
- *Satisfacción*. La comodidad y aceptación del modo de uso
- *Contexto de uso*. Los usuarios, objetivos, tareas, equipamiento (hardware, software y materiales) y el entorno físico y social en el que el producto es usado.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El proyecto se adecúa a las líneas prioritarias expuestas por la Facultad de Ingeniería de la UNJu en la Resolución FI N° 071/98, la cual incluye el área temática “Ingeniería de Software” en la cual se consideran las siguientes líneas de acción:

- Metodología de desarrollo de Software.
- Desarrollo ágil de software.
- Calidad de software.
- Verificación y Validación de software.
- Estándares para el desarrollo de software.

En la actualidad se trabaja en:

- El estudio de los atributos de calidad, en especial la norma ISO/IEC 25010.
- Comparación de trabajos científicos en los cuales se consideraron atributos de calidad en la construcción de las aplicaciones con RA.
- Comparación de metodologías o modelos de proceso que se utilizaron en aplicaciones con RA.
- Refinamiento de las fases de Meduc_AR de forma de incorporar a cada una de las fases los atributos de calidad pertinentes.

3. RESULTADOS OBTENIDOS/ESPERADOS

Para el presente proyecto, el cual tiene estipulado dos años de duración, se estableció como objetivo general:

- Realizar el refinamiento de la metodología Meduc_AR para el desarrollo de aplicaciones de Realidad Aumentada destinada al ámbito educativo.

Mientras que los objetivos particulares son:

- Analizar en profundidad las fases de la metodología Meduc_AR.
- Realizar estudios comparativos entre distintas metodologías o modelos de proceso utilizados en el desarrollo de aplicaciones de RA educativas, para determinar ausencias que permitan refinar Meduc_AR.
- Estudiar conceptos vinculados a las fases de la metodología, tales como elicitación de requerimientos, usabilidad, estándares de calidad para el desarrollo de software, verificación y validación de aplicaciones, entre otros.
- Realizar el refinamiento de la metodología, considerando los conceptos estudiados en el punto anterior.

Considerando los objetivos anteriores durante el año 2020 se obtuvo como resultado el trabajo “*Metodología para el desarrollo de aplicaciones con RA en Educación (Meduc_AR)*”. Cáceres, Nélica Raquel; Tolaba, Carolina; Gálvez Díaz, María del Pilar; Huerto Flores, Natalia María del; Hinojosa, Eduardo Elías. XXII Workshop de Investigadores en Ciencias de la Computación (WICC 2020, El Calafate, Santa Cruz). ISBN: 978-987-3714-82-5. Páginas: 430-434

Se logró avanzar en la evaluación de atributos de calidad a implementar en la metodología. Esta investigación también dio origen a la realización de una propuesta para trabajo final de grado de la carrera ingeniería informática. Se espera que los resultados de este trabajo contribuyan no solo al refinamiento de la metodología Meduc_AR sino también al área de investigación.

4. FORMACIÓN DE RECURSOS HUMANOS

El proyecto está siendo desarrollado por un equipo conformado por docentes investigadores de la Facultad de Ingeniería de la UNJu. La estructura del equipo de investigación es la siguiente:

- Directora: Mg. Ing. Nérida Raquel Cáceres. Categoría de Investigación IV. Investigadores:
- Ing. Ana Carolina Tolaba. Categoría de Investigación V. Actualmente realizando tesis de doctorado vinculada al área de modelado conceptual de datos a través de modelos semánticos.
- Mg. María del Pilar Gálvez Díaz. Categoría de Investigación III.
- Lic. Natalia María del Huerto Flores. Licenciada en Sistemas y alumna avanzada de la Carrera Ingeniería en Informática.

Los integrantes de este grupo de investigación participaron en:

- Curso de postgrado “*Realidad Aumentada*” dictado por la Universidad Nacional de la Plata, llevado a cabo los días 01 de octubre al 30 de noviembre de 2021 a cargo de la docente Dra. María José Abásolo, correspondiente al Doctorado en Ciencias Informáticas.
- Dirección de Trabajo Final de Carrera denominado “*Desarrollo de un prototipo con Realidad Aumentada para mostrar información de puntos de interés en la Facultad de Ingeniería de la UNJu aplicando la característica de usabilidad de ISO/IEC 25010*”. Alumno: Nicolás Ricardo Carrizo de la Carrera Ingeniería en Informática de la UNJu. Res. C.A.F.I. N° 511/21.

5. BIBLIOGRAFÍA

[1] Azuma, R. T. “A survey of augmented reality”. *Presence: Teleoperators and virtual environments*, 6(4), 1997, pp. 355-385.

[2] Gil, G. D., Arias Figueroa, D., Gimson Saravia, L. E., Sánchez, E., & Silvera, J. A. “Uso de realidad aumentada como complemento a los prácticos tradicionales de enseñanza-aprendizaje”. XVII Workshop de Investigadores en Ciencias de la Computación. Salta. 2015.

[3] Avison D. E. and Fitzgerald, G. *Information Systems “Development: Methodologies, Techniques, and Tools”*, McGraw-Hill. 1995.

[4] Ardila-Pérez, A. E. “Prototipo funcional para el uso de realidad aumentada en la EIA usando dispositivos móviles” *Doctoral dissertation, Administrativa, Financiera, Sistemas y Computación*. 2014.

[5] Tovar, L. C., Bohórquez, J. A., & Puello, P. “Propuesta metodológica para la construcción de objetos virtuales de aprendizaje basados en realidad aumentada”. *Formación universitaria*, 7(2), pp. 11-20. 2014.

[6] Solano, C. A., Díaz, J. F. C., & Bolaños, J. C. G. “Aplicación móvil de realidad aumentada para la enseñanza de la clasificación de los seres vivos a niños de tercer grado”. *Ingeniería*, 20(1). 2015.

[7] Abdulmushli, M. “Análisis de sistemas de realidad aumentada y metodología para el desarrollo de aplicaciones educativas”. *LSI1-Proyectos Fin de Máster. Universidad Rey Juan Carlos*. 2012.

[8] Cáceres, N. & Tolaba, A. “Metodología de Desarrollo para Aplicaciones con Realidad Aumentada en la Educación”. *V Congreso Nacional de Ingeniería en Informática/Sistemas de Información (CoNaIISI)*. 2017.

[9] ISO/IEC CD 25010 *Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Quality model and guide*, 2009.

[10] Fernández Zumaquero S. “Métodos de evaluación de la usabilidad para entornos de Realidad Virtual, Realidad Aumentada y Sistemas Ubicuos”, 2010.

Gobierno Digital. Evaluación del uso de la tecnología en el ámbito Universitario

Hugo Ramón¹, Nicolás Alonso², Emanuel Lazzari³, Augusto Villa Monte⁴, Sebastián Mir⁵, María Celeste Nobal⁶, Gustavo Iglesias⁷, Franco Bernaldo de Quirós⁷, Paola Santinelli⁷, César Collazos⁸,
 {hramon, nfalonso, elazzari, avillamonte}@docentes.unsada.edu.ar, {sebastian.mir, celeste.nobal}@unsada.edu.ar
 {giglesias, bfranco, psantinelli}@alumnos.unsada.edu.ar, ccollazos@unicauca.edu.co

Universidad Nacional de San Antonio de Areco (UNSAaA), Buenos Aires, Argentina
 Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA), Instituto de
 Investigación y Transferencia (ITT-CIC)

RESUMEN

Al *e-government*, se lo define desde dos perspectivas, una lo define como el uso de las TIC por agencias gubernamentales y otra más amplia lo define como un catalizador para inducir reformas administrativas y de política en el gobierno.

El objetivo de muchos países es tener un *e-government* con mayor presencia y cada vez más eficiente. Esto implica no solamente el uso de las TIC, sino una nueva interacción con los actores y el gobierno e involucra aspectos políticos y culturales, cambios tecnológicos y organizacionales, diseñados para direccionar un cambio profundo en las unidades de gobierno. Esta forma de interacción utiliza la tecnología con el fin de aumentar aspectos como la transparencia, eliminar las distancias y divisiones físicas y capacitar a los ciudadanos para participar en los procesos políticos, aplicando en diferentes dominios.

El *e-government* no significa automatizar viejas prácticas. El uso de computadoras en procesos complejos, puede lograr mejorar la calidad de los servicios entregados por el gobierno y/o promover la participación ciudadana, pero centrarse únicamente en las TIC no va a cambiar la mentalidad de los empleados, que es un aspecto central del éxito de la implementación de cualquier proyecto.

Este proyecto propone evaluar los esfuerzos y el impacto de las actividades de *e-government* en un ambiente educativo universitario, definiendo las características y factores críticos de su implementación, para luego poder realizar las recomendaciones que surjan de la evaluación de acuerdo a los lineamientos definidos en el Plan Nacional de Gobierno Electrónico (Decreto 378/2005 del PEN). En el caso de la Universidad Nacional de San Antonio de Areco (UNSAaA) se tomarán las normativas existentes y en caso de ser necesario se realizarán las propuestas necesarias que la complementen de manera de definir un proceso que permita la implementación del denominado "Gobierno Electrónico" en el ámbito de la UNSAaA.

Palabras clave: Gobierno Electrónico, Blockchain, Inteligencia Artificial, Ciencias de Datos.

CONTEXTO

Las líneas de investigación a describir se enmarcan en el proyecto de investigación: Gobierno digital. Evaluación del uso de la tecnología, con lugar de trabajo en la Universidad Nacional de San Antonio de Areco (UNSAaA) presentado en la convocatoria de Subsidios a la Investigación del año 2022 ante la Secretaría de Investigación de la Universidad.

¹ Doctorando, Magister, Lic. en Informática, Investigador ITT-UNNOBA-CIC, Profesor UNNOBA, UNSADA

² Maestrando, Lic. en Informática, Profesor UNNOBA, UNSADA

³ Ing. en Informática, Profesor UNNOBA, UNSADA

⁴ Dr. en Informática, Profesor UNSADA

⁵ Ing. en Informática, Profesor UNSADA

⁶ Lic. en Trabajo Social, Personal Administrativo UNSADA

⁷ Estudiante avanzado/a de la carrera de Lic. en Informática, UNSADA

⁸ Investigador externo

Su objetivo es evaluar las actividades de *e-government* planificadas y desarrolladas en el ámbito de la UNSaDA, y el impacto de las mismas en dicha institución.

1. INTRODUCCIÓN

Se define una tendencia tecnológica estratégica como aquella con potencial innovador sustancial que empieza a salir de un estado de emergente hacia un impacto de más amplio uso; o aquellas tecnologías que tienen rápido crecimiento, pero alta volatilidad y pueden alcanzar su punto de inflexión en los próximos años. La empresa experta en consultoría e investigación de las tecnologías de la información, *Gartner Inc.*, enunció que las tecnologías estratégicas de los últimos años se pueden resumir en: Tejido de datos, Malla de ciberseguridad, Computación de mejora de la privacidad, Plataformas nativas de la nube, Aplicaciones componibles, Inteligencia de decisiones, Hiperautomatización, Ingeniería de IA, Empresa distribuida, Experiencia total, Sistemas autónomos e Inteligencia artificial generativa.

Estas tecnologías y estratégicas impactan en cómo se desarrollan aplicaciones, buscando lograr la integración que las mismas deben implementar para descubrir las necesidades de los usuarios, con el fin de presentar la información pertinente en el lugar correcto y en el momento adecuado, con las características de calidad requeridas. La oportunidad de *I+D+i* se da en adaptar/diseñar aplicaciones o a la forma de construir las mismas, las cuales gestionan información procedente de redes de sensores compuestas por diversos nodos distribuidos y localizadas en espacios cerrados (hospitales, fábricas, oficinas, etc.) o abiertos (campos, bosque, etc.). Ambos espacios se encuentran en las diferentes actividades económicas de la región.

Además, la tecnología está incorporada en casi cada aspecto de nuestra vida, aunque en general pueda pasar desapercibida, está en nuestros teléfonos, en nuestros autos, en nuestros dispositivos de entretenimiento, en

nuestros electrodomésticos y en casi cualquier objeto de uso cotidiano, pequeñas computadoras capaces de procesar y almacenar información, las cuales no se perciben como objetos diferenciados, son ubicuos.

Dado estos nuevos sistemas de *software*, y teniendo en cuenta el contexto anteriormente definido, la calidad y la innovación juegan un papel muy importante en el desarrollo de productos y servicios. Esto implica que hoy en día incluir al usuario en los procesos de innovación, y ubicarlo en una posición centralizada, está dando lugar a la aparición de nuevos espacios de interacción y comunicación, en los que el actor que consume productos y servicios, deja de considerarse en forma pasiva para convertirse en un actor activo central en los procesos de *I+D+i*.

En Argentina, el Decreto 378/2005 del Poder Ejecutivo Nacional (PEN), indica que el Estado Argentino es el mayor productor/consumidor de información del país y establece el *Plan Nacional de Gobierno Electrónico*, el cual define los lineamientos estratégicos para *e-government*.

Es importante mencionar que el *e-government* no significa automatizar viejas prácticas. El uso de computadoras en procesos complejos, puede lograr mejorar la calidad de los servicios entregados por el gobierno y/o promover la participación ciudadana, pero centrarse únicamente en las TIC, no va a cambiar la mentalidad de los empleados, que es un aspecto central del éxito de la implementación de cualquier proyecto.

Los autores Al-Hashmi, A., & Darem, A. B. (2008), identifican diferentes estrategias para la implementar *e-government*:

- Trazar un plan integral a largo plazo.
- Identificar sólo unas pocas áreas clave como el foco de los primeros proyectos.
- Comenzar con proyectos pequeños en fases.

Para la elaboración de planes e iniciativas, algunos autores dividen el proceso de aplicación en diversas fases. Estas fases no son dependientes entre sí, ni consecutivas, por lo cual no necesitan completar una fase antes de iniciar otra. Ofrecen maneras de pensar acerca de los objetivos.

En Baum, C., & Di Maio, A. (2000), se describe un proceso de cuatro fases, que poseen diferentes objetivos y requerimientos relacionados con la capacitación en el uso de las TIC y costos asociados.

Es un objetivo del presente proyecto evaluar los esfuerzos y el impacto de las actividades de *e-government* en un ambiente educativo universitario. A los fines indicados, se aplicarán las denominadas Tecnologías de la Información y de la Comunicación (TIC) a la organización y a los procedimientos internos de la administración en red. En este contexto, toma relevancia el concepto de *Gobierno Inteligente*, el cual refiere a la incorporación masiva de las TIC para mejorar los procesos internos gubernamentales que impactan en la interacción con las personas, organizaciones y la entrega de servicios. Para ello se debe de desarrollar, mantener y promover los sistemas integrados basados en Internet para la prestación de servicios y la provisión de información; y adoptar las medidas necesarias para que las comunicaciones se efectúen preferentemente mediante tecnologías informáticas, optimizando para ello la utilización de los recursos electrónicos disponibles en las distintas dependencias de la Universidad.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las líneas de investigación de este trabajo consisten en el estudio de casos, de tipo descriptivo-correlacional, que mezcla técnicas cualitativas que van desde entrevistas, observación, revisión de documentos, observación directa de las instalaciones, de los procesos y de los productos que aplican al objeto de estudio.

El estudio de casos es una de las diversas maneras de investigar que dependen de tres condiciones:

- El tipo de pregunta de investigación.
- El control que un investigador tiene sobre la conducta actual de los eventos.
- El foco sobre lo contemporáneo como opuesto al fenómeno histórico.

Como se indica en Yin, R. (1994), en general los estudios de casos son las estrategias preferidas cuando se realizan las preguntas “cómo” y “porqué” y cuando el foco está en un fenómeno contemporáneo dentro de un contexto de la vida real. En Muñiz M. (2010), se indica que los estudios de caso son importantes en la investigación cualitativa, pero al mismo tiempo se usan cada vez más en una serie de aplicaciones en el campo profesional por ejemplo en la evaluación de un sistema educativo, la organización de una empresa, la participación social en un grupo, etc. Todo esto desde la perspectiva de las personas específicas que forman parte de esos grupos, empresas o instituciones.

La metodología de investigación aplicada que se utilizará en este proyecto consta de las siguientes etapas:

- Analizar el estado actual y los fundamentos teóricos de los temas propuestos.
- Realizar seminarios de discusión internos y externos para identificar problemas tecnológicos. Es importante el intercambio de información con los grupos del país y del exterior que trabajan en proyectos similares.
- Recopilar y analizar la bibliografía existente, así como también, publicaciones realizadas por otros equipos de investigación, revistas y demás producciones digitales disponibles en la Web sobre la temática de estudio.

- Focalizar el objetivo de investigación descriptiva para conocer situaciones, costumbres y actitudes predominantes a través de la descripción exacta de las actividades, objetos, procesos y personas.
- Investigar las alternativas de solución buscando optimizar el empleo de recursos humanos y tecnológicos interactuando con el mundo real obteniendo una solución específica para el problema concreto. Esta etapa resulta particularmente importante en la formación de recursos humanos que se inician en *I+D+i* y es beneficioso el intercambio con otros grupos con proyectos de investigación similares.
- Integrar el trabajo al equipo de investigación en el marco del proyecto al que pertenece.
- Analizar hipótesis, experimentar y obtener resultados e indicadores.
- Evaluar los resultados, eventualmente publicarlos y realizar transferencia al medio.
- Publicar resultados obtenidos.

3. RESULTADOS OBTENIDOS/ESPERADOS

Se espera que los aspectos investigados contribuyan a propiciar el fortalecimiento en la formación de recursos humanos, en su rol de investigadores o participantes activos en equipos de investigación; fomentando la culminación de sus estudios superiores, promoviendo la redacción, exposición y defensa de trabajos finales de grado y postgrado.

A su vez, se pretende difundir y transferir los logros alcanzados mediante la presentación y participación en diferentes congresos, jornadas y workshops de carácter nacional e internacional relacionados con el uso de las tecnologías emergentes.

En lo que refiere a la evaluación y caracterización de las actividades de *e-government*, se espera que los aspectos investigados contribuyan a poder optar por un enfoque cuanti-cualitativo que identifique y describa un estado de situación acerca de las actividades *e-government* en el caso de UNSAdA, que permite luego hacer un mapeo de las decisiones de política institucional que inciden en el *e-government* o el valor público de este. Desde la perspectiva cuantitativa, se espera que, por medio del diseño de un instrumento de medición, cuyos objetivos alcancen cuestiones de infraestructura, prácticas e iniciativas, políticas de decisión institucional respecto al *e-government*, permita la identificación de tendencias acerca de las políticas y de los usos de los servicios entregados por el *e-government* en el contexto de un relevamiento no exhaustivo que contribuya un muestreo valioso para una primera sistematización.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está compuesto por docentes e investigadores formados y en formación pertenecientes a la Universidad Nacional de San Antonio de Areco (UNSAdA), algunos de los cuales dirigieron becas del Consejo Interuniversitario Nacional: (CIN) y de la Comisión de Investigaciones Científicas (CIC) y diversos trabajos finales de la carrera de Analista en Informática. Entre los integrantes del equipo hay un Doctor y un Magíster, y otros docentes investigadores se encuentran cursando estudios de postgrados. Se espera en el 2022 tener dos presentaciones de becas CIN de estudiantes avanzados y una beca CIC.

5. BIBLIOGRAFÍA

Al-Hashmi, A., & Darem, A. B. (2008). Understanding phases of E-government project. *New Delhi. Disponible en: http://www.csi-sigegov.org/emerging_pdf/17_152-157.pdf*.

Baum, C., & Di Maio, A. (2000). Gartner's four phases of e-government model. *Gartner Group*, 12.

Decreto 378/2005 (2005). Plan Nacional de Gobierno Electrónico y Planes Sectoriales de Gobierno Electrónico. Disponible en: <http://servicios.infoleg.gob.ar/infolegInternet/anexos/105000-109999/105829/norma.htm>

Decreto 996/2018 (2018). Agenda Digital Argentina. Disponible en: <http://servicios.infoleg.gob.ar/infolegInternet/verNorma.do?id=316036>

Fath-Allah, A., Cheikhi, L., Al-Qutaish, R. E., & Idri, A. (2014). E-government maturity models: A comparative study. *International Journal of Software Engineering & Applications*, 5(3), 71.

Gobierno Digital del Departamento de Estado de los EEUU (2012). Digital Government Strategy - United States Department of State. Disponible en: <https://www.state.gov/digital-government-strategy/>

Kraemer, K. L., & King, J. L. (1984). National Policies for Local Government Computing: an Assessment of Experience in Ten OECD Countries. *International Review of Administrative Sciences*, 50(2), 133-147. DOI: <https://doi.org/10.1177/002085238405000204>

Krieger, M. (2009). Los desafíos de cambio en la Administración Pública Argentina. *Schweinheim, Guillermo (compilador), Estado y administración pública: críticas, enfoques y prácticas en la Argentina actual*, Buenos Aires, Asociación Argentina de Estudios en Administración Pública, pág, 71-93.

Layne, K., & Lee, J. (2001). Developing fully functional E-government: A four stage model. *Government information quarterly*, 18(2), 122-136.

Muñiz, M. (2010). Estudios de caso en la investigación cualitativa. *División de estudios de posgrado en la Universidad Autónoma de Nuevo León. Facultad de psicología. México*, 1-8.

Ronzhyn, A., Spitzer, V., & Wimmer, M. (2019). Scenario Technique to Elicit Research and Training Needs in Digital Government Employing Disruptive Technologies. 20Th Annual International Conference On Digital Government Research On - Dg.O 2019. DOI: [10.1145/3325112.3325231](https://doi.org/10.1145/3325112.3325231)

Serale, F., Redl, C., & Munte, A. (2019). Blockchain en la administración pública: ¿Mucho ruido y pocos bloques? *Inter-American Development Bank*. DOI: <https://doi.org/10.18235/0001951>

Tellechea, T. (2018). El gobierno electrónico como derecho y la brecha digital en Argentina. *Informe Integrar*.

Turing, S. (2012). *Alan M. Turing: Centenary Edition*. Cambridge University Press.

Valdés G., Solar, M., Astudillo, H., Iribaren, M., Concha, G., & Visconti, M. (2011). Conception, development and implementation of an e-government maturity model in public agencies. *Government Information Quarterly*, 28(2), 176-187.

ESTUDIO PRELIMINAR PARA EL DESARROLLO DE UNA HERRAMIENTA DE BÚSQUEDA EN REPOSITORIOS ACADÉMICOS

María del Pilar Gálvez, Sergio L. Martínez, Nélide R. Cáceres, Ana C. Tolaba, José R. Quispe, Felipe F. Mullicundo, Laura R. Villarrubia

Facultad de Ingeniería - Universidad Nacional de Jujuy
Ítalo Palanca 20 San Salvador de Jujuy – 0388 4221576
nrcaceres@fi.unju.edu.ar

RESUMEN

A través de repositorios digitales de Acceso Abierto, instituciones académicas buscan exponer su producción científica/académica. No obstante, los datos publicados pueden resultar insuficientes ya sea porque no cuentan con la terminología adecuada para su descripción o bien los metadatos empleados para la descripción de los datos publicados, al ser semiestructurados no permiten explotar la información de mejor manera porque hay conocimiento implícito que favorece la descripción de nuevas relaciones entre los datos explicitados que no está siendo usado. Todo esto limita la realización de búsquedas más integrales y eficaces de forma de obtener mejores resultados.

Ante esta situación, el presente proyecto plantea en primer lugar, el estudio de repositorios digitales que utilizan las instituciones académicas con el fin de definir todos los conceptos relacionados al mismo y que sean adecuados para el repositorio de la Facultad de Ingeniería, para definir estrategias de búsquedas, técnicas y herramientas que incorporen tecnologías de web semántica y sistemas NoSQL. Este artículo presenta el estado de avance alcanzado en este proyecto, los resultados y la formación de recursos humanos concretada en el marco del mismo.

Palabras clave: *NoSQL, Repositorios Digitales, Semántica, Búsquedas Mejoradas.*

CONTEXTO

La línea de investigación que se presenta en este trabajo corresponde al proyecto “*Desarrollo de Herramienta de Búsqueda utilizando Web Semántica y Sistemas NoSQL*” aprobado por la Secretaría de Ciencia y Técnica de la Universidad Nacional de Jujuy. Este proyecto es categoría A (código D/0168) y se encuentra bajo incentivo.

1. INTRODUCCIÓN

La tendencia de exponer la producción de las instituciones a través de repositorios digitales de Acceso Abierto [1] ha incrementado la cantidad de repositorios digitales notablemente. El Ministerio de Ciencia y Tecnología, en Argentina, impulsó la creación de repositorios de Acceso Abierto mediante el Sistema Nacional de Repositorios Digitales en C y T (SNRD). Además, elaboró un proyecto de ley para “Creación de repositorios Digitales Institucionales de Acceso Abierto, Propios o Compartidos” que fue aprobado a fines de 2013. La ley 26899 establece la obligatoriedad del acceso abierto a la producción financiada con fondos públicos a nivel nacional a través de repositorios digitales que las instituciones deberán crear, mantener e integrar al SNRD [2].

La implementación de un repositorio digital ofrece diferentes beneficios tanto para investigadores, estudiantes, así como al resto de la sociedad ya que permiten crear y compartir conocimiento, y facilitan la

transferencia de conocimiento al sector productivo [3]. Para lograr estos beneficios es necesario considerar la organización de la información disponible en los repositorios de forma tal que el conocimiento implícito que favorece la descripción de nuevas relaciones entre los datos explicitados sea aprovechado.

En este sentido, es necesario que un repositorio cuente con metadatos precisos, completos y con un formato homogéneo, esto le permitirá interoperar con otros repositorios para realizar intercambio de información además de crear servicios de valor añadido [4].

La web semántica, es otra tecnología a ser considerada para obtener los beneficios indicados anteriormente. La web semántica permite el acceso inteligente y preciso a grandes repositorios de datos, favoreciendo a la difusión del contenido de los repositorios [5]. Dentro de las tecnologías de la Web semántica se dispone de RDF (Resource Description Framework) que permiten dotar de significado los datos y transacciones de datos en la Web [6]. También ofrece ontologías, estas son estructuras más completas que permiten una representación formal de un concepto, además de la representación semántica y sintáctica del mismo [7].

Estas tecnologías permitirán la recuperación de información mediante búsquedas semánticas. Las cuales se refieren a una búsqueda de conceptos no solo basada por la comparación de palabras (búsqueda sintáctica), sino por deducciones lógicas que consideran la intención y el significado contextual de las palabras empleadas en la búsqueda [8].

La implementación de una herramienta para la realización de búsquedas semánticas comprende el empleo de modelos de datos que incluyen información semántica la cual puede ser gestionada mediante sistemas NoSQL [9], [10]. El objetivo de los modelos de datos semánticos es capturar el significado de los datos mediante la integración de conceptos relacionales con conceptos de abstracción más poderosos.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El proyecto se adecúa a las líneas prioritarias expuestas por la Facultad de Ingeniería de la UNJu en la Resolución FI N° 071/98, la cual incluye el área temática “Ingeniería de Software”, en la cual se consideran las siguientes líneas de acción: Repositorios digitales, Gestión de la información y el conocimiento, Sistemas de información web y bases de datos y Recuperación de la información. En la actualidad se trabaja en:

- Estudio de los repositorios digitales implementados en la actualidad para definir las características del repositorio a realizar, como así también el buscador adecuado.
- Definición de los metadatos para los trabajos finales de grado y la estrategia para la recopilación de los mismos.
- Definición de las características del repositorio digital de la Facultad de Ingeniería con información de trabajos finales de grado.
- Definición de la estrategia de búsqueda a desarrollar utilizando web semántica y bases de datos NoSQL.

3. RESULTADOS OBTENIDOS/ESPERADOS

Este proyecto tiene estipulado cuatro años de duración, y se establecieron los siguientes objetivos.

Como objetivo general, el proyecto de investigación tiene como propósito desarrollar una herramienta de búsqueda que facilite el análisis y comprensión de los datos almacenados en el repositorio digital de trabajos finales de grado de la Facultad de Ingeniería de la UNJu. La herramienta propuesta combinará para su desarrollo, tecnologías de web semántica y sistemas NoSQL. La información extraída de estos repositorios será utilizada como apoyo para la toma de decisiones, tanto a nivel

administrativo y operativo de los estudiantes de grado ya que les proporciona el conocimiento necesario para llevar a cabo la selección del tema de trabajo final. Además, esta información permitirá que otros usuarios como egresados, docentes, investigadores y agentes externos conozcan las diferentes líneas de investigación de los trabajos desarrollados, logrando de esta forma la transferencia de la UNJu hacia la comunidad.

Además se establecieron como objetivos particulares:

- Realizar un estudio de los repositorios digitales.
- Realizar un estudio de web semántica y sistemas NoSQL.
- Efectuar un análisis respecto de los metadatos y la forma de acceder al contenido de los repositorios institucionales, por ejemplo, análisis de motores de búsqueda empleados.
- Ejecutar pruebas mediante distintos tipos de consultas, con los datos de repositorios digitales institucionales de libre acceso, que permitan realizar un análisis de los resultados obtenidos.
- Realizar un estudio de la implementación de búsquedas semánticas en bases de datos NoSQL.
- Definir el tipo de bases de datos NoSQL y los motores de búsqueda adecuados para el repositorio digital propuesto.
- Realizar el relevamiento de los datos de los proyectos finales de grado de la Facultad de Ingeniería de la UNJu.
- Analizar la estructura de metadatos de los proyectos finales relevados.
- Desarrollar una herramienta de búsqueda mediante la combinación de web semántica y bases de datos NoSQL.
- Generar distintos tipos de búsquedas utilizando la herramienta desarrollada.
- Evaluar los resultados obtenidos por la herramienta de búsqueda a través de pruebas de aceptación.
- Comparar los resultados obtenidos tanto por la herramienta propuesta como por el sistema actual de consulta SIBUNJU.

Considerando los objetivos descritos anteriormente durante el año 2020 se obtuvo como resultado el trabajo *“Herramienta de búsqueda en repositorios académicos basada en web semántica y sistemas NoSQL”*. Gálvez Días, María del Pilar; Martínez, Sergio L.; Cáceres, Nélica R.; Tolaba, Ana C.; Villarrubia, Laura R.; Mullicundo, Felipe F.; Quispe, José R.; Sanguero Ballon, Marcelo R.; Sandoval, Iván L.; Quispe, Jairo J.M.; Lamas, Daniel A., XXII Workshop de Investigadores en Ciencias de la Computación (WICC 2020, El Calafate, Santa Cruz). ISBN: 978-987-3714-82-5. Páginas: 425-429.

Mediante la realización de un curso de posgrado se avanzó en la identificación del modelo de metadatos para el repositorio propuesto para la facultad de ingeniería. Se prevé continuar con la incorporación de web semántica para el desarrollo de una herramienta que búsqueda que permita mejorar los resultados de búsquedas exhaustivas de antecedentes sobre trabajos concluidos en la unidad académica a la cual pertenece para dotar a su trabajo final de originalidad.

4. FORMACIÓN DE RECURSOS HUMANOS

El proyecto está siendo desarrollado por un equipo conformado por docentes investigadores del Grupo de Investigación y Desarrollo en Ingeniería de Software (GIDIS) de la Facultad de Ingeniería de la Universidad Nacional de Jujuy. La estructura del equipo de investigación es la siguiente:

- Directora: Mg. María del Pilar Gálvez. Categoría de Investigación III.
- Codirector: Mg. Ing. Sergio Luis Martínez. Categoría de Investigación III.

Investigadores:

- Mg. Ing. Nélica Raquel Cáceres. Categoría de Investigación IV.
- Ing. Ana Carolina Tolaba. Categoría de Investigación V. Actualmente realizando tesis de doctorado vinculada al área de modelado conceptual de datos a través de modelos semánticos.

- Esp. Ing. Laura Rita Villarrubia. Categoría de Investigación IV.
- Lic. Felipe Fernando Mullicundo. Categoría de Investigación V.
- Mg. Ing. José Rolando Quispe.

Con la realización de este proyecto de investigación se espera la consolidación de los miembros del grupo, además de la formación de jóvenes investigadores principalmente alumnos avanzados de las carreras afines de la facultad de ingeniería. Se espera formar nuevos trabajos finales de grado y participación en becas, cuyas temáticas serán propias del mencionado proyecto de investigación.

Los integrantes de este proyecto de investigación participaron en:

- Curso de postgrado “*Bibliotecas y Repositorios Digitales. Tecnologías y Aplicaciones*” dictado por la Universidad Nacional de la Plata, llevado a cabo los días 1 al 5 de marzo de 2021, a cargo de la docente Dra. Marisa De Giusti, correspondiente al Doctorado en Ciencias Informáticas.
- Curso de capacitación denominado “*Administrador de base de datos no relacionales*” dictado por la Fundación Carlos Slim en modalidad online. Disponible en <https://capacitateparaeempleo.org/pages.php?r=.tema&tagID=4066>
- Dirección de Trabajo Final de Carrera denominado “*Coopertino: Aplicación web que asiste a la formación de grupos de estudio y el trabajo en equipo*”. Alumnos: Eduardo Andrés Albornoz y Matías Ramón Ruiz de la Carrera Ingeniería en Informática de la UNJu. Res. FI N° 352/2021.

5. BIBLIOGRAFÍA

- [1] Maenza, R., & Darin, S. (2016). Universidades abiertas trabajando en la innovación tecnológica y la transparencia. *Revista Internacional Transparencia e Integridad, RITI* nro, 2.
- [2] Peña, K. I. C. (2014). Modelos de acceso abierto en educación y ciencia. *Educación y educadores*, 17(2), 8. DOI: 10.5294/edu.2014.17.2.7
- [3] Ramírez, M. R., Soto, M. D. C. S., Moreno, H. B. R., Rojas, E. M., Millán, N. D. C. O., & Cisneros, R. F. R. (2019). Metodología SCRUM y desarrollo de Repositorio Digital. *Revista Ibérica de Sistemas e Tecnologías de Informação, (E17)*, 1062-1072.
- [4] Delgado, J. C. S., & Alvarado, M. A. C. (2017). Repositorios institucionales digitales: Análisis comparativo entre SEDICI (Argentina) y Kérwá (Costa Rica). *e-Ciencias de la Información*, 1-32.
- [5] Sulé, A., Centelles, M., Franganillo, J., & Gascón, J. (2016). Aplicación del modelo de datos RDF en las colecciones digitales de bibliotecas, archivos y museos de España. *Revista española de documentación científica*, 39(1), 121.
- [6] McBride, B. (2004). Theresource description framework (RDF) and its vocabulary description language RDFS. In *Handbook on ontologies* (pp. 51-65). Springer, Berlin, Heidelberg.
- [7] Gruber, T., Ontology, I. L. L., & Özsu, M. T. (2009). *Encyclopedia of database systems*. Springer-Verlag, ISBN 978-0-387-49616-0.
- [8] Portolés Sánchez, M. J. (2010). *Búsqueda semántica en repositorios de conceptos biomédicos estandarizados: CT Hunter* (Doctoral dissertation).
- [9] Venkatraman, S., Fahd, K., Kaspí, S., & Venkatraman, R. (2016). SQL Versus NoSQL movement with big data analytics. *Int. J. Inform. Technol. Comput. Sci*, 8, 59-66.
- [10] “NOSQL Databases,” Disponible en: <http://nosql-database.org> Acceso: Octubre, 2019.

Agrupamiento de Escenarios guiado por objetivos

Gladys N. Kaplan¹, Jorge H. Doorn^{2,3}, Candela Santander¹

¹Departamento de Ingeniería e Investigaciones Tecnológicas. Universidad Nacional de la Matanza. Argentina

²Escuela de Informática, Universidad Nacional del Oeste

³Departamento de Ingeniería, Universidad Nacional de Tres de Febrero
gkaplan@unlam.edu.ar, jdoorn@uno.edu.ar

RESUMEN

Los Escenarios son usados en primer lugar para describir el contexto donde se planifica insertar el sistema de software y luego para describir como se espera que se desempeñe ese contexto con el sistema en servicio. Tanto en el caso del contexto pre-existente, como en el del caso de contexto planificado, cuando los Escenarios están completos y consistentes, se agrupan jerárquicamente en Escenarios Integradores. Este mecanismo permite reducir las omisiones y facilita la comprensión global del contexto en estudio. Sin embargo, cuando este contexto debe ser analizado desde una perspectiva particular, como por ejemplo durante la validación, el agrupamiento jerárquico falla. Esto ha sido un factor determinante para que algunas actividades del proceso de requisitos basado en Escenarios no alcancen aún la calidad esperada. Se propone, en el presente artículo, el agrupamiento de los Escenarios según objetivos específicos, que tome información de todos los Escenarios y ofrezca una visualización adaptada a cada circunstancia. Esto facilita, entre otras actividades, la verificación y la validación. Se ha comprobado inicialmente, que este agrupamiento también puede mejorar el uso de los Escenarios por los desarrolladores y reducir la curva de aprendizaje en la capacitación.

Palabras clave: Proceso de Requisitos, Escenarios, Agrupamiento por objetivos, Visualización.

CONTEXTO

La línea de investigación que se presenta es parte del proyecto de investigación “Agrupamiento selectivo de Escenarios” de la Universidad Nacional de La Matanza.

1. INTRODUCCIÓN

Este proyecto está enmarcado en la Ingeniería de Requisitos (IR) [1] [2] [3], en un proceso de requisitos particular [4] cuyos modelos utilizan el lenguaje natural (LN). Su objetivo es alcanzar una profunda comprensión del contexto antes de definir el sistema de software. Dicha estrategia se compone de dos grandes etapas bien distinguibles: una de aprendizaje y la otra de definición. De existir conocimiento previo, la primera etapa, se convierte en una actividad confirmatoria. Los modelos que se utilizan en este proceso son:

- Un modelo léxico, LEL (Léxico Extendido del Lenguaje) [5] [6], el cual describe el léxico del contexto.
- Un modelo organizacional que describe el proceso del negocio tal como existe al momento del comenzar la IR, denominado Escenarios Actuales [7].
- Un modelo organizacional que describe el proceso del negocio con el sistema de software incluido [8] [9], denominado Escenarios Futuros.
- Un modelo de requisitos donde se especifican los requisitos del software.

Los Escenarios futuros tienen empotrados los requisitos del software. Por lo tanto, el documento de requisitos (ERS) se obtiene extrayendo los requisitos del software desde estos Escenarios [10].

Los modelos organizacionales se componen de un conjunto de Escenarios que describen el contexto donde se desempeñará el futuro sistema de software, desde el punto de vista de los propios actores de ese macrosistema: los clientes y los usuarios. Cada Escenario

describe una situación del contexto. Se construyen dos conjuntos de Escenarios, el primero describe el macrosistema observable, mientras que el segundo describe un posible macrosistema futuro planificado, con el sistema de software en funcionamiento. Cada uno de estos conjuntos de Escenarios permite comprender los detalles particulares, pero no son aptos para obtener una visión global del macrosistema. Por este motivo se construye un pequeño grupo de Escenarios Integradores que procuran atemperar esta dificultad. Los Escenarios Integradores están organizados jerárquicamente, según el flujo de trabajo. Existen algunos casos donde los Escenarios ordenados de esta manera dificultan la percepción deseada, por ejemplo, cuando se desea analizar el rol de un determinado usuario o la evolución de un producto. Para analizar este fenómeno se describe el caso particular de la validación, la cual ha sido estudiada en proyectos previos [11] [12]. Para comprender la validación de Escenarios es necesario destacar la diferencia entre la validación de los Escenarios Actuales y la de los Escenarios Futuros. En el primer conjunto se encuentran descritas las situaciones del contexto actual, las cuales son de amplio conocimiento de los clientes y usuarios. En este caso la validación es más sencilla para el cliente o usuario que para los ingenieros/as de requisitos ya que son estos los que deben afrontar todos los problemas de comprensión involucrados. Por el contrario, en el caso de los Escenarios Futuros, la dificultad se invierte ya que estos Escenarios describen un plan del proceso de negocio futuro que no es observable. La comprensión del contexto futuro requiere de un esfuerzo muy importante, principalmente para los clientes o usuarios. En este caso son los ingenieros/as de requisitos quienes están mejor posicionados. Es de esperar que en cada caso sea necesario analizar los Escenarios desde perspectivas particulares, según el momento del proceso de requisitos y de las necesidades individuales. Para lograrlo, es necesario seleccionar la información de los Escenarios, la cual se obtiene agrupándolos por

un objetivo común. Cabe destacar que además de agruparlos es necesario contar con una visualización que muestre solo aquella información que responda al objetivo buscado sin perder contextualización.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Los modelos y actividades del proceso de requisitos utilizados en este proyecto, propenden a mantener al usuario integrado al mismo, con el mínimo esfuerzo posible. Para concretar esta aspiración se construye el modelo LEL y el modelo Escenarios, ambos en lenguaje natural con un pequeño formalismo. El único agrupamiento de Escenarios existente en el proceso de requisitos es a través de los Escenarios Integradores, donde se seleccionan los Escenarios involucrados en bloques jerárquicos, los cuales son luego analizados por el equipo de desarrollo o con el cliente-usuario. Pero se ha detectado que cuando son utilizados para analizar aspectos puntuales del contexto, estos Integradores pierden efectividad. Esto se debe a diferentes factores, como ser la cantidad de Escenarios, la complejidad del contexto, la dispersión de la información, etc. Se debe recordar que el proceso de requisitos toma relevancia para sistemas de software de mediana y gran envergadura donde la cantidad de Escenarios es un dato significativo. Además, el uso del lenguaje natural tiende a dificultar la comprensión cuando el volumen de información es grande. Por tal motivo, no solo es necesario contar con otros tipos de agrupamiento para contextos específicos sino también una visualización adecuada de la información que se obtenga del agrupamiento.

En el presente proyecto se propone una nueva forma de agrupamiento de Escenarios, en este caso según un *objetivo*, permitiendo mostrar versiones resumidas de los Escenarios en grupos con un objetivo común. Algunos ejemplos de *objetivos* pueden ser conocer “qué información contiene el formulario de solicitud”, “quién y cuándo se completa el

formulario de solicitud”, “en qué situaciones se rechaza el formulario de solicitud”, etc.

Este agrupamiento no solo selecciona los Escenarios pertinentes, sino que además cruza la información con los documentos del contexto, convirtiéndose en una verificación. Entre los documentos mencionados se encuentra la literatura organizacional (manuales de procedimientos, protocolos, etc.) y las transcripciones de las entrevistas realizadas durante la IR hasta ese momento.

Hasta el momento el mecanismo de agrupamiento propuesto está basado en el análisis de los símbolos del LEL relacionados a partir de un conjunto de símbolos semillas, pero se espera encontrar otras formas de agrupamiento de Escenarios como la utilización de Orientación a Aspectos [13] [14], prioridades, orden de implementación, seguridad, etc.

Se espera que los avances obtenidos puedan ser utilizados también para capacitación con aprendizaje autónomo, pero a su vez selectivo, impactando positivamente en el costo del proyecto. También se espera que el agrupamiento impacte en otras actividades del proceso de construcción del software, como el diseño y la codificación mejorando la comprensión de información que provee la IR.

3. RESULTADOS OBTENIDOS/ESPERADOS

Con el objetivo de corroborar la hipótesis acerca de que el agrupamiento por objetivos mejora la comprensión de los Escenarios, se realizaron diferentes pruebas con tres casos de estudio, los resultados obtenidos fueron muy promisorios.

El primer paso fue definir el objetivo del agrupamiento. Luego, buscar Escenarios que estén relacionados con el objetivo del agrupamiento. De esta manera, primero se obtienen los Escenarios relacionados con el objetivo y luego, se repite el mismo procedimiento en los documentos del contexto. Finalmente, con esta información se construyó la visualización correspondiente utilizando una versión resumida del modelo Escenarios. Las

pruebas se realizaron en los casos *Obtención de Pasaporte, Control de Calidad, Norpak y Homologaciones*.

Con la información obtenida en las pruebas se construyó una heurística preliminar que se describe a continuación, la cual ha sido refinada en función a la experiencia de los participantes en el proceso de la construcción de Escenarios en más de un centenar de casos.

A continuación, se describen los pasos de la heurística preliminar:

1) Extraer palabras clave desde el Objetivo

- *Identificar los símbolos del LEL íntimamente relacionado con el objetivo del agrupamiento.*
- *Buscar en ellos los símbolos relacionados*
- *Algunas nuevas palabras clave se van obteniendo al analizar los episodios. O sea, se obtienen nuevas palabras clave por proximidad léxica.*

Objetivo: <i>Dónde participa la OP Urgente</i>			
Escenario	Ubicación Geog.	Precond.	EPISODIOS
GENERAR ORDEN DE PRODUCCIÓN URGENTE	Oficina de planificación	La orden de producción debe tener un plazo de entrega menor a 10 días.	1. Si la orden de producción presenta dificultades para complementarse ENTONCES el oficial planificador transforma la orden de producción a orden de producción urgente .
EMITIR PROGRAMA DE FABRICACIÓN EXTRAORDINARIO	Oficina de planificación	Que exista una orden de producción urgente cuyo plazo de entrega sea menor a 5 días. Que durante un fin de semana no se pueda cumplir con una orden de producción urgente	2. El encargado de planta comunica a la oficina de planificación que no se puede cumplir el programa de fabricación con una orden de producción urgente .
GENERAR INFORME DE DIFICULTADES DE FABRICACIÓN	Oficina de planificación	Debe existir un programa de fabricación que contenga una orden de producción urgente	3. El encargado de planta evalúa si se puede cumplir el programa de fabricación con una orden de producción urgente . 4. Si no se puede cumplir con el programa de fabricación ENTONCES el encargado de planta debe informarlo utilizando un formulario "Informe de Dificultades de Fabricación"

Tabla 1 – Información para la visualización

2) Buscar y extraer información de los Escenarios

- Utilizar el Escenario integrador para respetar el orden.
- Analizar si existe relación entre el Escenario y el Objetivo del agrupamiento. Es suficiente con un solo episodio que cumpla con este ítem.

En este punto se crea una tabla con la información para la visualización. En algunos casos solo se copiará un solo episodio mientras que, en otros, será un conjunto o todos. Completar la Tabla 1 con la información del Escenario. Para ejemplificar las tablas se utilizó el Caso Norpak que es una fábrica de Cajas de Cartón. Esto permitió un mejor cruce de la información y permitió analizar la heurística al tener varios conjuntos de Escenarios de diferentes autores.

3) Buscar palabras clave en los documentos

En este caso se vuelven a utilizar las palabras clave.

Repetir para cada documento existente

Repetir para cada palabra clave

- Seguir el rastro de la palabra clave desde el LEL al texto
- Asegurarse que el párrafo esté relacionado con el Objetivo del agrupamiento.
- Marcar cada párrafo seleccionado con un identificador, por ejemplo, como en la Fig. 1
- Eliminar conceptos duplicados.

Se debe prestar una especial atención principalmente en las transcripciones de las entrevistas, ya que la información suele estar muy dispersa y donde la coincidencia de contenidos es más difícil de detectar.

Objetivo: "Dónde participa la OP Urgente"
a) Las cajas se entregan a los 30 días corridos de recibirse la orden de compra, aunque se ofrece un descuento especial por órdenes recibidas con una anticipación de 60 días y se admiten órdenes urgentes con una anticipación de 10 días, pero a las mismas se les aplica un sobrepeso importante.
b) Esta política comercial está basada en la necesidad de planificar la producción en forma apropiada.
c) Las ordenes urgentes tienen asociado el riesgo de utilizar demasiada materia prima, provocando un descarte de material excesivo, con el natural incremento en los costos de producción.
d) La planificación de la producción se realiza priorizando aquellas ordenes de producción que ocasionen el menor desperdicio de materias primas.
e) Sin embargo, si alguna orden de producción, difícil de complementar con alguna otra, se atrasa en demasía y el plazo de entrega se reduce a menos de 10 días, la misma se transforma en una orden de producción urgente.
f) Las ordenes de producción urgentes son siempre incluidas en los programas de fabricación más próximos posibles.
g)

Fig.1 – Texto del Caso Norpak

4) Crear tabla de coincidencia

En este momento, se tiene la tabla con la información de los Escenarios (ver Tabla 1) y el texto con los párrafos identificados. Es momento de relacionarlos, como se muestra en la Tabla 2.

- Para cada episodio, buscar si existe alguna parte del texto equivalente.
- Si se encontró, marcar con una tilde en el documento, para saber que fue incorporado.
- Si se encontró texto, completar la Tabla 2.
- Analizar el texto para identificar otras palabras clave.

5) Detectar información faltante en los Escenarios

En este punto se está en una verificación, ya que lo que se busca es determinar omisiones. De esta manera, como ya se mencionó, para Validar es necesario asegurar la completitud y consistencia de la información. Por lo tanto, la preparación de la información para Validar se transforma en un mecanismo de verificación.

- Analizar si los párrafos no incluidos son omisiones en los Escenarios.
- Buscar en el texto aquellos párrafos que no fueron incluidos, aquellos que no tienen tilde y analizar la pertinencia según el Objetivo.
- Si es una omisión, crear una lista con los párrafos que deben ser analizados antes de incorporar la información en los Escenarios correspondientes.

Objetivo: Donde participa la OP Urgente	
Texto	Episodios
a) y e)	1)
b), c), d), f), g), J)	-----
h)	3) y 4)
i)	2)

Tabla 2 – Tabla de Coincidencia

6) Completar la tabla de visualización

Se debe entregar a los autores de los Escenarios la información omitida para que modifiquen los Escenarios, si corresponde. Luego, se debe completar la Tabla 1.

Una vez identificadas las discrepancias, errores y omisiones en los Escenarios se deben impactar los cambios correspondientes y

repetir el proceso para completar ambas tablas.

Cabe destacar que algunos problemas se podrán corregir inmediatamente y otros, se convertirán en dudas para una futura validación. En este caso se puede observar que los párrafos b), c), d), f), g) y J) no han sido incorporados oportunamente en los Escenarios y los mismos están incompletos. La cantidad de defectos dependerá de la calidad de los Escenarios construidos.

Cabe destacar que es posible encontrar información en los Escenarios que no se corresponde con ningún texto, esto se debe a que puede existir información obtenida de manera informal, como es una charla en un pasillo o tomando un café donde aparece información relevante que es incorporada a los Escenarios.

Para el futuro, se espera probar otras técnicas de agrupamiento y encontrar la mejor forma de visualizar la información obtenida.

4. FORMACIÓN DE RECURSOS HUMANOS

La línea de investigación presentada es parte directa de la tesis doctoral de la Mg. Gladys Kaplan. También es parte de la formación inicial en investigación de la alumna Candela Santander.

5. BIBLIOGRAFÍA

- [1] Davis, A. M (1993), *Software Requirements: Objects, Functions and States*, Prentice Hall, Englewood Cliffs, NJ.
- [2] Elizabeth Hull, Ken Jackson, Jeremy Dick (2010), *“Requirements Engineering”*, Springer, Third Edition, ISBN-10: 1849964041, ISBN-13: 978-1849964043.
- [3] Pohl, K. (2010), *“Requirements Engineering: fundamentals, principles, and techniques”*, Springer Publishing Company, Incorporated.
- [4] Leite, J., C., S., P., Doorn, J. H., Kaplan, G., Hadad, G., D., S., Ridao, M., N., (2004) *“Defining System Context using Scenarios”*, in *Perspectives on Software Requirements*, Kluwer Academic Publishers, cap.8, pp. 169-199.
- [5] Leite J.C.S.P., Franco, A.P.M., (1990) *“O Uso de Hipertexto na Elicitação de Linguagens da Aplicação”*, Anais de IV Simpósio Brasileiro de Engenharia de Software, SBC, pp. 134-149.
- [6] Hadad, G.D.S., Doorn J.H., Kaplan G.N. (2008), *“Creating Software System Context Glossaries”*, in *Encyclopedia of Information Science and Technology*, Second edition, Mehdi Khosrow-Pour (ed.), Idea Publishing.
- [7] Leite, J. C. S., Hadad, G. D. S., Doorn, J. H., & Kaplan, G. N. (2000), *“A scenario construction process”*, *Requirements Engineering Journal*, Springer-Verlag London Ltd., Vol.5, N°1, pp. 38-61.
- [8] Doorn J.H., Hadad G.D.S., Kaplan G.N. (2002), *“Comprendiendo el Universo de Discurso Futuro”*, WER'02: Workshop de Engenharia de Requisitos, Valencia-España.
- [9] Kaplan, G.N., Doorn, J.H., Gigante, N.C. (2013), *“Evolución Semántica de Glosarios en los Procesos de Requisitos”*, XIX Congreso Argentino de Ciencias de la Computación, CACIC 13, Mar del Plata.
- [10] Hadad, G., Doorn J., Kaplan G. (2009) *“Explicitar Requisitos de Software usando Escenarios”*, WER'09: Workshop de Engenharia de Requisitos, Medellín, Colombia.
- [11] Kaplan Gladys N., Doorn Jorge H., Hadad Graciela D.S. (2003) *“Validación de Escenarios Futuros con prototipos”*, anales de WICC 2003 – Workshop de Investigadores en Ciencias de la Computación – UNICEN - Tandil –, Argentina, 22 y 23 de mayo de 2003
- [12] Kaplan, G.N., Doorn, J.H., Guatelli R., Gigante N., Hadad, G.D.S.; (2009), *“Storyboard Basados en Escenarios Futuros”*, anales de WICC 2009 – XI Workshop de Investigadores en Ciencias de la Computación – UNSJ - San Juan, Argentina, 7 y 8 de mayo 2009, pp 306-309; ISBN 978-950-605-570-7.
- [13] E. Baniassad and S. Clarke. (2004) *Theme: An approach for aspect oriented analysis and design*. En *International Conference on Software Engineering*, 2004.
- [14] I. Brito, (2004) *Aspect-Oriented Requirements Engineering*. Proceeding of the 7th International Conference on Unified Modelling Language (UML).

Utilización de los Modelos de Procesos en los Procesos de Requisitos

Gladys N. Kaplan¹ y Gabriel E. Blanco¹

¹Departamento de Ingeniería e Investigaciones Tecnológicas. Universidad Nacional de la Matanza. Argentina
gkaplan@unlam.edu.ar, gblanco@unlam.edu.ar

RESUMEN

Las organizaciones determinan su desempeño a través de una cadena de procesos. Contar con información tangible y concreta de cada proceso es una forma de evitar fallos y estar preparados para los cambios. Ante contextos altamente competitivos, las empresas se ven obligadas a enfocarse cada vez más en sus procesos y dejar lentamente la tradicional verticalidad. La definición de los procesos está fuertemente alineada a la organización, asegurando una perspectiva adecuada para entender el contexto en estudio. Además, estos modelos son comprendidos por todos los usuarios y están consensuados por ellos. En el presente artículo se describe un mecanismo para utilizar los modelos de procesos durante la Ingeniería de Requisitos, particularmente para aquellas estrategias que estudian el contexto actual antes de definir los requisitos del software. Se ha comprobado que esta incorporación reduce el tiempo de modelado y mejora la cooperación de los usuarios al eliminar la necesidad de aprender nuevos modelos. Además, se evita duplicar información. También se ha detectado que la relación entre los modelos de procesos y los modelos de requisitos permiten una validación cruzada, ya que la omisión o inconsistencias en un modelo es rápidamente detectada al construir el otro.

Palabras clave: Proceso de requisitos, modelos de procesos, organizaciones enfocadas en los procesos.

CONTEXTO

La línea de investigación que se presenta “Utilización de los Modelos de Procesos en los Procesos de Requisitos” se encuentra dentro

del proyecto “Agrupamiento selectivo de Escenarios” de la Universidad Nacional de La Matanza.

1. INTRODUCCIÓN

Los nuevos desafíos a los que se enfrentan las organizaciones requieren que las mismas conozcan cada vez mejor sus procesos. Mercados muy cambiantes y clientes cada vez más exigentes obligan a las empresas a buscar métodos que las ayuden a funcionar a su máxima capacidad. En este sentido, las organizaciones actuales se están orientando a un *enfoque por procesos* [1] [2] para tener mayor control y poder adaptarse rápidamente a los cambios. Los beneficios de este enfoque son:

- Mejor aprovechamiento de los recursos.
- Mayor facilidad en la toma de decisiones.
- Eliminación de actividades que no aportan ningún valor a la empresa.
- Reducción del tiempo de operaciones.
- Clientes satisfechos.

Estas organizaciones tienen los procesos definidos, documentados y consensuados por los usuarios. Según la ISO 9001-2015 [3], conocer los procesos proporcionan múltiples beneficios:

- Aumento de la capacidad de centrar los esfuerzos en los procesos clave y en las oportunidades de mejora.
- Resultados coherentes y previsibles mediante un sistema de procesos alineados.
- Optimización del desempeño mediante la gestión eficaz del proceso, el uso eficiente de los recursos y la reducción de las barreras interdisciplinarias.
- Posibilidad de que la organización

proporcione confianza a las partes interesadas en lo relativo a su coherencia, eficacia y eficiencia

La incorporación de este enfoque al estándar advierte un futuro cercano donde las organizaciones que aún no han comenzado a definir sus procesos comiencen a hacerlo, proyectando una mayor frecuencia de esta documentación, la cual es de suma importancia para la Ingeniería de Requisitos (IR) [4] [5] [6].

En el presente proyecto se ha trabajado particularmente con un proceso de requisitos [7] el cual tiene una estrategia de tres etapas: Comprender el UdeD¹ actual, Proyectar el UdeD futuro y Explicitar los Requisitos del Software (ver Figura 1). La primera etapa propone conocer el dominio en estudio antes de generar una propuesta para el nuevo sistema de software. Para ello se elicit información del dominio y se modela el proceso de negocio tal como existe al comenzar la IR. Es en esta etapa donde la documentación de los procesos se hace más relevante. El conocimiento obtenido en esta primera etapa es utilizado para la segunda, Proyectar el UdeD futuro, donde se toman las decisiones acerca de los servicios que tendrá el nuevo sistema de software. La complejidad de esta etapa se debe a la necesidad de proyectar cómo será el proceso del negocio con el sistema de software incluido. Para ello se modelan todas las situaciones involucradas con el nuevo sistema de software, siendo estos modelos los anfitriones de los requisitos. Finalmente, en la tercera etapa, estos requisitos son extraídos y explicitados en un documento. El formato de este documento dependerá de las políticas organizacionales y de los estándares nacionales o internacionales que se utilicen. Durante todo el proceso se utilizan básicamente dos modelos: el Léxico Extendido del Lenguaje (LEL) [10] [11] y los Escenarios [8]. El LEL es un glosario cuyo objetivo es

describir el léxico del dominio para mejorar la comunicación con el cliente y asegurar la comprensión de todos los artefactos producidos. Los Escenarios son narrativas estructuradas de situaciones del contexto, centrandó la atención en el comportamiento. Estos modelos pueden representar diferentes perspectivas dependiendo del momento en el cual se construyen. El LEL es el glosario del UdeD actual y evoluciona al LEL_R [12] en el UdeD futuro. Lo mismo sucede con los Escenarios, los cuales representan el UdeD actual denominados Escenarios actuales (EA) o el UdeD futuro, Escenarios futuros (EF). Los Escenarios se relacionan internamente jerárquicamente. Cuando una situación más pequeña está contenida en otra, aparece un sub-escenario. Cuando un escenario tiene una mirada global del contexto es un Escenario Integrador, en este caso nuevamente puede ser actual (EAI) o futuro (EFI). Los integradores se construyen cuando los Escenarios están completos, verificados y validados.

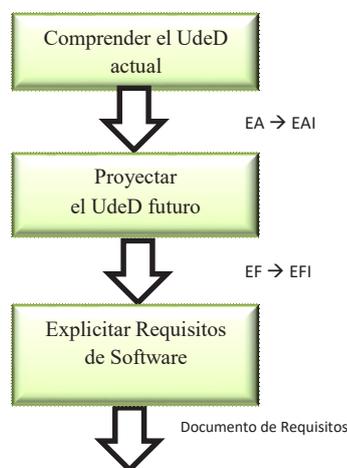


Figura 1 – Estrategia de la IR

Si bien la IR de utiliza la literatura del contexto (manuales, protocolos, etc.), dicha información suele tener algún grado de dispersión con respecto a los modelos de

¹ UdeD: "Todo el contexto en el cual el software será desarrollado y operado. Incluye todas las fuentes de información y todas las personas relacionadas con el software. Se utiliza el término Universo de

Discurso con el mismo significado que lo utiliza Michael Jackson en [13].

requisitos que se desean construir. Si bien el mismo proceso de requisitos debe neutralizar esta dificultad, existe el riesgo de omitir información relevante que se pierde en la dispersión. Además, estos documentos organizacionales suelen ser generados tomando el punto de vista del “deber ser”, o sea, cómo la organización espera que se realicen las tareas, pudiendo aparecer contradicción con la información operativa observable. Esto obliga al ingeniero/a de requisitos a modelar ambos puntos de vista y luego, cuando se deban definir los servicios del nuevo sistema de software, tomar una decisión. La ventaja de los modelos de procesos es que estos puntos de vista están unificados, ya que en la definición de cada proceso participan diferentes usuarios y se los define por consenso.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Este proyecto es la continuación de varios proyectos de investigación previos donde se ha estudiado el Proceso de Requisitos basado en Escenarios [7]. Desde su creación, uno de sus objetivos ha sido minimizar el esfuerzo del usuario para comprender los modelos construidos. Por tal motivo, se han utilizado representaciones cercanas a los usuarios o que requieran un mínimo esfuerzo de aprendizaje, basadas en la mayoría de los casos, en *lenguaje natural*. De esta manera se ha mejorado la participación y cooperación del usuario durante toda la IR. Pero por mínimo que sea, debe aprender cómo interpretar los modelos de requisitos para comprender el proceso. En la propuesta de la presente línea de investigación se espera dar un salto más al utilizar los mismos modelos organizacionales, anulando el esfuerzo mencionado. Estos modelos son construidos por los mismos usuarios. Utilizar estos modelos permite:

- Obtener requisitos de software más alineados a la organización.
- Mejorar la comunicación entre clientes, usuarios y desarrolladores.

- Eliminar información duplicada entre los modelos organizacionales y los de requisitos.
- Mejorar la propuesta para el nuevo sistema de software ya que el Enfoque por Procesos se enmarcan en la Gestión de Calidad.

Cabe mencionar que algunas organizaciones no han comenzado aún a definir sus procesos y las que lo han hecho, presentan diferente grado de madurez en estas definiciones [10]. El presente proyecto se concentra en aquellas empresas que ya han comenzado en este camino, siendo algunos de los modelos utilizados aquellos que tienen una mirada global del negocio como los Mapas de Procesos [3] y los que describen el despliegue o analizan diferentes aspectos de los procesos como SIPOC (Suppliers, Inputs, Process, Output, Customers), TMAP (Thought Process Map), diagramas de flujo de procesos, etc.

Se ha podido comprobar que entre los modelos de procesos y los del proceso de requisitos existe mucha información común. Por un lado, entre los mapas de procesos y los Escenarios Integradores y luego, entre los de despliegue de procesos y los Escenarios. En este contexto, el presente proyecto estudia la inclusión de los modelos de procesos en el Proceso de Requisitos basado en Escenarios, específicamente durante la etapa ***Comprender el UdeD actual***. Se espera probar la estrategia en organizaciones con diferente grado de madurez en el enfoque por procesos.

Cabe mencionar que el LEL se construye en todos los casos y que se sugiere incorporarlo a la documentación organizacional. En este caso se espera analizar si la construcción del LEL también se verá alterada al modificar el proceso.

Si bien esta línea de investigación está concentrada en la etapa ***Comprender el UdeD actual***, se espera avanzar en la etapa ***Proyectar el UdeD futuro***, donde se construyen los requisitos del software. En esta etapa aún no se ha trabajado. En este caso el cambio de estrategia es radical, generando posiblemente una “IR enfocada en los procesos”. Aún se desconoce de qué manera los modelos de procesos pueden

albergar los requisitos del software, especialmente los requisitos no funcionales (RNF). También, se espera avanzar en una estrategia de retroalimentación entre los modelos de requisitos y los de procesos, generando un mapa de procesos “futuro” con los procesos involucrados con el sistema de software en ejecución para generar los modelos de procesos automáticamente.

1. RESULTADOS OBTENIDOS/ESPERADOS

Para probar el cubrimiento de los modelos de procesos durante la Comprensión del UdeD actual, se realizaron diferentes pruebas con el caso Norpak², para el cual existen los modelos LEL y Escenarios. En este caso se construyeron los modelos de procesos, se comenzó por el mapa de proceso y luego se construyeron los SIPOC correspondientes. Se ha podido observar en diferentes aproximaciones a otras organizaciones, que inicialmente existen dos grupos. En el primer grupo están aquellas organizaciones que solo tienen la definición general de los procesos, específicamente el mapa de procesos. Este caso es aplicable a todas las organizaciones que han comenzado a orientarse a los procesos, pero algunas en particular, solo tienen este mapa. Para Norpak se utilizó la información del mapa como guía para construir los EA (ver Figura 2-1). Se construyó el Escenario General (EG) a partir del mapa de procesos, suprimiendo así la necesidad de identificar las situaciones en el contexto, lo que ha provocado algunos problemas de completitud en los Escenarios. Para construir el EG se copió el nombre de cada proceso como episodio del EG. El resto de los componentes del EG se completaron en función al Objetivo General del Sistema. Luego, se buscó información en el contexto para construir cada EA.

Es interesante mencionar que el proceso de requisitos en sus primeras versiones proponía la construcción de Escenarios con una estrategia Top-Down, donde se construía un

Escenario General y a partir de él se describían los EA. Este EG no era seguro ya que su origen era la elicitación de situaciones del contexto, sea con observación, entrevistas, documentación. Esto provocaba que a medida que se conocía el contexto se lo comprendía mejor y el EG debía ser actualizado. Este esfuerzo adicional de mantener el EG tenía además un alto riesgo de cometer desvíos significativos y no detectarlos hasta avanzado el proceso. Esto fue determinante para cambiar la estrategia a un enfoque Bottom-Up, construyendo todos los EA e integrando al final. Por tal motivo, poder construir un EG desde información segura, permite retomar las ventajas de un enfoque Top-Down, pero ahora, sin riesgos.

El segundo grupo, corresponde a organizaciones que han avanzado en la definición de los procesos. En estos casos, se cuenta con el mapa de procesos y el despliegue de los mismos (ver Figura 2-2). En el análisis a Norpak se analizó el UdeD actual con los diagramas organizacionales. Se pudo observar que los modelos de procesos reemplazaron satisfactoriamente a los EA, reduciendo significativamente el tiempo de modelado de toda la primera etapa del proceso. Se espera en el futuro medir los tiempos y el nivel de comprensión con cada estrategia para compararlas. También, fue necesario analizar el cubrimiento de las definiciones de procesos con respecto al Objetivo General de Sistema. O sea, determinar si el cubrimiento de las definiciones de los procesos era total o parcial en relación a los procesos involucrados en la IR. En este caso fue total. Una vez identificados los procesos se separaron para analizar el conjunto. En la comparación de la información de procesos y de las descripciones de los EA, se identificaron dos tareas que no habían sido vistas por los EA y otra fue descripta como parte de una tarea mayor, minimizando su relevancia.

² Empresa que fabrica Cajas de Cartón Corrugado a medida.

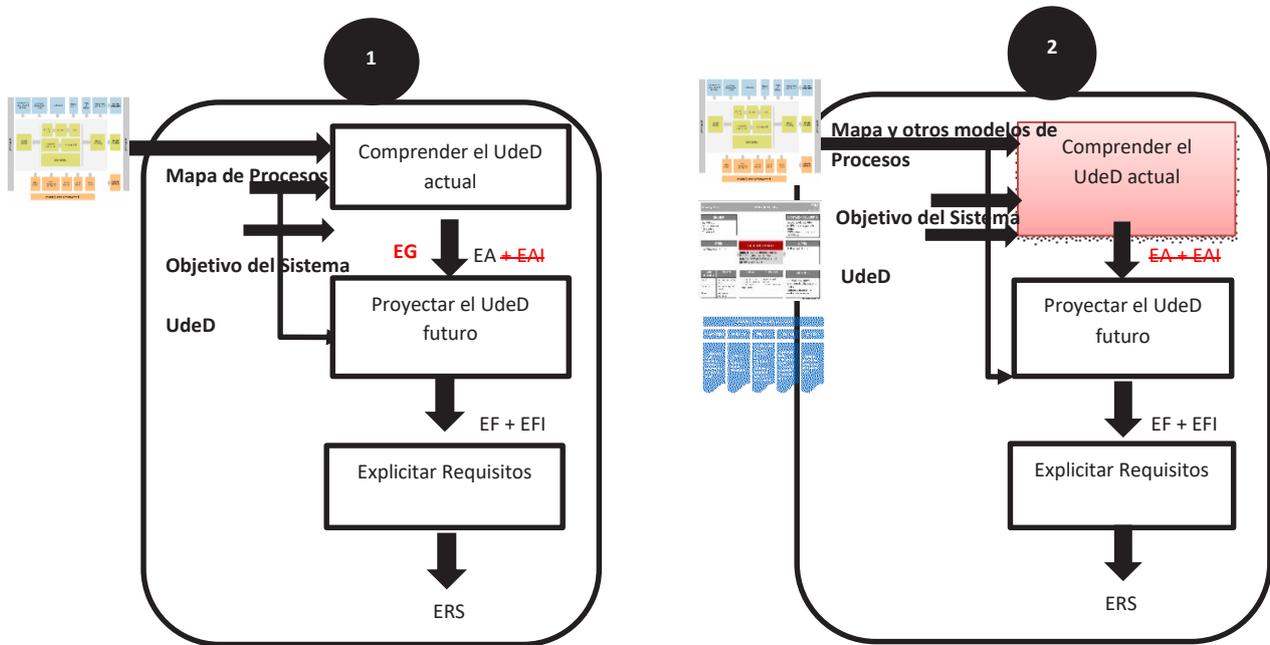


Figura 2 – Estrategias de utilización de los Modelos de Procesos en el Proceso de Requisitos

FORMACIÓN DE RECURSOS HUMANOS

La línea de investigación presentada es parte directa de las tesis doctorales de la Mg. Gladys Kaplan y del Mg. Gabriel Blanco.

5. BIBLIOGRAFÍA

- [1] Félix González Benítez, (2014), “Enfoque por procesos, mejora de los procesos”, Editorial Académica Española, ISBN 3659048992, 9783659048999
- [2] R.G. Lee , B.G. Dale, (1998), “Business process management: a review and evaluation”, Business Process Management Journal, ISSN: 1463-7154.
- [3] “Sistemas de gestión de la calidad – Requisitos” (2015) (ISO 9001-2015). <https://www.iso.org/obp/ui/#iso:std:iso:9001:ed-5:v1:es>
- [4] Davis, A. M (1993), “Software Requirements: Objects, Functions and States”, Prentice Hall, Englewood Cliffs, NJ.
- [5] Elizabeth Hull, Ken Jackson, Jeremy Dick (2010), “Requirements Engineering”, Springer, Third Edition, ISBN-10: 1849964041, ISBN-13: 978-1849964043.
- [6] Pohl, K. (2010), “Requirements Engineering: fundamentals, principles, and techniques”, Springer Publishing Company, Incorporated.
- [7] Leite, J., C., S., P., Doorn, J. H., Kaplan, G., Hadad, G., D., S., Ridao, M., N., (2004) “Defining System Context using Scenarios”, in Perspectives on Software Requirements, Kluwer Academic Publishers, cap.8, pp. 169-199.
- [8] Leite Julio, Hadad Graciela, Doorn Jorge Horacio, Kaplan Gladys, (2000), “A Scenario Construction Process”, Requirements Eng 5:38-61 2000 Springer-Verlag London Limited.
- [9] Jan vom Brocke and Michael Rosemann, (2015), “Handbook on Business Process Management”, editores, Springer Link, ISBN: 978-3-642-45100-3.
- [10] Leite, J.C.S.P., (1989) “Application Languages: A Product of Requirements Analysis”, Computer Science Department of PUC-Rio, Brazil.
- [11] Hadad, G.D.S., Doorn, J.H., (2007), “Creating Software System Context Glossaries”, Encyclopedia of Information Science and Technology, Idea Group Publishing, 2º edición.
- [12] Kaplan G.N, Doorn J.H., Gigante N., (2013) “Evolución Semántica de Glosarios en los Procesos de Requisitos”. CACIC 2013.
- [13] Jackson, M., “The Role of Architecture in Requirements Engineering”, (1994), ICRE’94, First International Conference on Requirements Engineering, IEEE Computer Society, Colorado Springs, p.241.

Evaluación de calidad en datos abiertos. Mejora de procesos. Sistemas resilientes en la gobernanza digital.

Silvia Esponda , Ariel Pasini , Marcos Boracchia ,
Rocío Muñoz , Juan Ignacio Torres , Patricia Pesado 

Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata
50 y 120 La Plata Buenos Aires
Centro Asociado CIC
526 e/ 10 y 11 La Plata Buenos Aires

(sesponda, apasini, marcosb, rmunoz, jitorres, ppesado) @lidi.info.unlp.edu.ar

RESUMEN

El III-LIDI (Instituto de Investigación en Informática LIDI) posee una línea de investigación orientada a la calidad de software, mejora de procesos y gobernanza digital.

Esta línea abarca temas relacionados con la calidad de software, en particular la evaluación de calidad de producto y calidad de datos y la viabilidad de la aplicación de los estándares internacionales ISO/IEC en bases de datos abiertas. Por otro lado, la línea abarca conceptos relacionados con mejora de procesos, gobernanza digital y sistemas resilientes.

Palabras Claves

Ingeniería de Software – Calidad del proceso – Calidad del producto – Calidad de Datos - Normas de Calidad – ISO - Gobernanza Digital

CONTEXTO

Esta línea de investigación se enmarca en el proyecto “11/F023 Metodologías, técnicas y herramientas de Ingeniería de Software en escenarios híbridos”. Mejora de proceso” y en el subproyecto “Gobernanza Digital. Mejora de Procesos” (2018-2022), acreditado por el Ministerio de Educación de la Nación.

Además, la línea se encuentra abordada por el proyecto “Calidad de Datos” aprobado por la Facultad de Informática UNLP.

El Instituto posee diversos acuerdos de cooperación con varias Universidades de Argentina y del exterior y con empresas privadas del sector, interesadas en mejorar sus procesos de desarrollo aplicando mejoras y participa en proyectos internacionales, entre ellos ERASMUS “Fortalecimiento de la capacidad de gobernanza de ciudades sostenibles inteligentes”.

1. INTRODUCCION

En los últimos años, es cada vez más notoria la importancia de los datos en una organización. El avance digital está influyendo en todos los sectores y ha convertido a los datos en el recurso más potente y en un aspecto clave para la toma de decisiones. Sin embargo, estos sectores no suelen disponer de recursos accesibles que evalúen la calidad de sus datos. Los datos se pueden ver afectados por factores negativos: ruido, valores perdidos, inconsistencias, un tamaño demasiado grande en cualquier dimensión, entre otros.

Los datos abiertos son aquellos datos accesibles y reutilizables, sin exigencia de permisos específicos. El crecimiento del

volumen de este tipo de datos es generado a gran velocidad y en forma continua.

El recurso encontrado en los datos abiertos ha demostrado el poder que posee la ciudadanía teniendo la información correcta y actualizada en la toma de decisiones.

Una baja calidad de estos datos puede generar información insatisfactoria y resultados que no son de utilidad. Es por ese motivo que uno de los puntos del proyecto, apunta a analizar y aplicar los estándares de calidad de datos, propuestos por la familia ISO/IEC 25000 (SQuaRE) en particular las normas ISO/IEC 25012 - "Data Quality Model" [1], la cual define un conjunto de características destinadas a evaluar la calidad de los datos, e ISO/IEC 25040 - "Evaluation process" [2], que define el proceso de evaluación a llevar a cabo.

En una ciudad inteligente sostenible (CIS), los ciudadanos constantemente generan información que luego será utilizada en la toma de decisiones sobre el desarrollo de esa ciudad y, pasado un tiempo, influirá en la vida de esos ciudadanos.

Lograr poner a disposición de la sociedad datos de interés común y con calidad puede dar como resultado el desarrollo de una idea o aplicación innovadora que proporcione nuevos datos, conocimientos u otros servicios. Además, el hecho que los ciudadanos tengan acceso a los datos y se les permita participar en su análisis para construir el desarrollo de la ciudad, es un importante aporte para la ciudad inteligente. [3]

Según [4] las ciudades inteligentes sostenibles representan las últimas etapas de progresión a través de ciudades digitales y ciudades inteligentes, consideradas como un proceso transformador continuo, basado en la colaboración y el compromiso de diferentes actores, construyendo diferentes capacidades (humanas, técnicas e institucionales) de manera de mejorar la

calidad de vida, proteger los recursos naturales y perseguir el desarrollo socioeconómico. La Unión Internacional de Telecomunicaciones (UIT) de la ONU, estableció una de las primeras definiciones de ciudad inteligente sostenible: "Una ciudad inteligente sostenible es una ciudad innovadora que utiliza tecnologías de información y comunicación (TIC), y otros medios, para mejorar la calidad de vida, la eficiencia de la operación y los servicios urbanos y la competitividad, al tiempo que se garantiza que satisfaga las necesidades de las generaciones, presentes y futuras, con respecto a los aspectos económicos, sociales, ambientales y culturales".

Las CIS se componen de cinco pilares: *Social, Económico, Ambiental, Gobernanza e Infraestructura Urbana*, para los que se requiere la prestación de servicios públicos que satisfagan las necesidades de sus ciudadanos [5].

La pandemia hizo visible el concepto Ciudad Digital Resiliente. La transformación digital ha tomado fuerza como detonante de la resiliencia, en un proceso acelerado por la pandemia que apresuró cambios (que ya se evidenciaban) en las interacciones, la navegación de espacios y el modo de acceder a los servicios.

Estos cambios se han dado además, pues en las décadas venideras se espera que el proceso de urbanización en Latinoamérica sea prácticamente total [6]. La población se desplaza hacia las ciudades en busca de oportunidades comerciales y económicas. Frente a estos problemas, las ciudades se enfrentan al riesgo de padecer impactos y tensiones.

A fin de prepararse para resistir con el menor trastorno posible y prevenir, minimizar y gestionar estos escenarios de forma efectiva, las ciudades deben hacerse más resilientes.

La transformación digital es una herramienta clave de la resiliencia para prepararse para los desastres y las

catástrofes y para mejorar el involucramiento ciudadano. [7]

Un sistema resiliente debe tener las siguientes cualidades:

1. Flexible: Refuerza la voluntad y capacidad para adoptar estrategias alternativas ante los cambios de circunstancias y las crisis súbitas
2. Redundante: Genera capacidad adicional para contemplar alteraciones imprevistas, facilitando así múltiples vías para satisfacer cualquier objetivo necesario
3. Integrado: Incorpora una variedad de sistemas e instituciones que permiten catalizar beneficios adicionales, mediante el uso compartido de los recursos
4. Inclusivo: Da prioridad a participación y consultas amplias para generar un sentido de responsabilidad compartido en la toma de decisiones
5. Ingenioso: Reconoce usos alternativos de los recursos para alcanzar los objetivos establecidos, especialmente en tiempo de crisis
6. Robusto: Desarrolla sistemas bien planteados para garantizar que cualquier fallo sea previsible, seguro, y proporcional a su causa
7. Reflexivo: Recurre a la experiencia para definir decisiones futuras y modificar estándares establecidos con anterioridad

Tener presentes estas cualidades permite observar el funcionamiento de sus ciudades desde una perspectiva de resiliencia y les da la capacidad de diseñar estrategias más adecuadas para su crecimiento urbano

En este caso desde el proyecto se busca evaluar las aplicaciones para establecer si poseen las cualidades definidas. Ofreciendo pautas, recomendaciones y sugerencias para la mejora de las mismas.

Otras de las líneas del proyecto relacionadas con las CIS son los servicios de voto por internet, facilitando el proceso de elecciones a diferentes organismos públicos y privados.

La definición de CIS incluye el concepto de “*innovadora*”. En este sentido desde el proyecto se analizan diferentes técnicas y metodologías de innovación y co-creación que les permiten a los ciudadanos contribuir a la resolución de problemas públicos, como por ejemplo Ciencia de datos, Ciencia del comportamiento, Simulación y modelaje, Inteligencia colectiva y Design thinking. [8]

Además, el grupo participa del proyecto CAP4CITY “Fortalecimiento de la capacidad de gobierno para ciudades inteligentes y sostenibles” del Programa ERASMUS. [9]

En dicho proyecto se ha creado una metodología que guía y orienta el diseño de MOOC, aplicada para el diseño de 31 cursos. En particular se ha desarrollado los cursos de “Aplicaciones para Ciudades Inteligentes” y “Diseño y Co-Creación de Servicios Orientados al Usuario” y estamos trabajando junto a la Universidad Nacional del Sur en la definición de una Maestría en Ciudades Inteligentes Sostenibles.

Dentro de la línea de la investigación del proyecto, se destacan los siguientes ejes principales:

1- Calidad de datos.

Actualmente se está trabajando en el diseño de una propuesta que facilite la aplicación de los estándares nombrados anteriormente para la evaluación y análisis de la calidad de los datos de organismos públicos y privados.

2- Mejora de los servicios de gobierno digital en organismos públicos de gobierno.

Se continúa trabajando en la generación de portales institucionales y además se está desarrollando una herramienta para la vinculación de datos abiertos.

Desde el año 2003, el III-LIDI trabaja en aplicaciones relacionadas con la gobernanza digital para la elección de

autoridades mediante el voto electrónico. Se destacan la definición e implementación de tres modelos (presencial, semipresencial y remoto) en distintos tipos de votaciones (urnas electrónicas, ambientes de votación, comunicaciones, entre otras).

3- Mejora en los procesos de gestión de la Facultad de Informática

El III-LIDI, en coordinación con el área de Gestión Electrónica Digital y Calidad de la Facultad de Informática, trabaja en este proyecto, con el objeto de analizar, definir y establecer un plan a ser aplicado a distintos procesos de la Gestión Universitaria.

La Facultad de Informática posee un Portal de Gestión Administrativa que provee a sus Alumnos, Docentes, No Docentes y Graduados una guía para la realización de trámites de los principales servicios que presta. De cada uno de estos trámites se brinda información estandarizada sobre: Descripción, Forma de Acceso, Requisitos y Contacto [10]

Desde el año 2011, la Facultad de Informática ha iniciado el camino hacia la certificación de distintos procesos.

El curso de Nivelación a Distancia logró en el año 2012 la certificación IRAM-ISO 9001:2008 del Sistema de Gestión de Calidad (SGC) del "Diseño y realización del curso de Nivelación a Distancia para el Pre-Ingreso a la Facultad de Informática". En 2018 se realizó la migración de la certificación a la versión IRAM-ISO 9001:2015 y se actualizó el alcance del SGC.

Asimismo, en 2016 se ha obtenido la certificación IRAM-ISO 9001:2008 del SGC del área de Concursos Docentes: "Llamado a Concursos Docentes Ordinarios", "Concurso Auxiliar Docente Ordinario" y "Concurso Profesor Ordinario" de acuerdo a la ordenanza 179 de la Universidad Nacional de La Plata, y las ordenanzas 303 y 308 de la Facultad de Informática. En 2018 se realizó la

migración de la certificación a la versión IRAM-ISO 9001:2015.

En el marco de la certificación Euro-Inf, la Facultad ha iniciado el proceso de implementación un Sistema Interno de la Calidad (SIGCFI) que abarca los cuatro pilares de la Universidad: formación, investigación, extensión y transferencia

4- Ciudades digitales resilientes.

Desde el proyecto se busca generar patrones para evaluar y definir la concordancia de las app con las cualidades establecidas para que una app sea resiliente.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

- Análisis y estudio de normas y modelos orientados a la Calidad de Datos.
- Aplicación de los conceptos de Calidad de Datos a la base de datos abierta de los casos de COVID publicada por los organismos gubernamentales.
- Asistencia en el proceso de certificación en organismos públicos y privados según los requisitos de IRAM - ISO 9001.
- Desarrollo de Voto electrónico presencial y Voto por Internet. Arquitecturas adaptadas a la legislación vigente.
- Análisis, discusión y estudio de herramientas para la aplicación de normas de calidad relacionadas con certificación de servicios gubernamentales.
- Análisis, discusión y estudio de técnicas y herramientas de mejoras de proceso en el desarrollo de software.
- Análisis de propuestas de innovación y co-creación de servicios públicos
- Análisis de aplicaciones móviles para ciudades digitales
- Análisis de resiliencia en sistemas para la Gobernanza digital

3. RESULTADOS OBTENIDOS/ ESPERADOS

- Se avanza en el desarrollo de una herramienta para asistir en la evaluación de calidad de datos utilizando GQM
- Se avanza en el proceso de recertificación bajo la norma IRAM-ISO 9001:2015: “Concursos Docentes de la Facultad de Informática”.
- Se analizan nuevos procesos de la Facultad, con posibilidad de generar guías/instructivos para facilitar la prestación de los servicios y analizar la posibilidad de su certificación.
- Se mantiene el Portal de Gestión Administrativa para la Facultad de Informática UNLP, donde se reúne información y enlaces sobre los diferentes servicios que brinda la facultad.
- Se realizaron acciones de consultoría y asesoramiento en organismos públicos y privados.
- Se presentó el informe final de la tesis de doctorado “Modelo de madurez de los servicios de gobierno electrónico en el ámbito universitario”.
- Se aplicó el Voto por Internet en distintos organismos.
- Se avanza en la implementación del SIGCFI de la Facultad.

4. FORMACION DE RECURSOS HUMANOS

- El proyecto cuenta con becarios de Maestría de la UNLP en los temas del área.
- Se desarrollan tesis de doctorado y tesinas de grado en el área.
- Se dirigen dos tesinas de grado en temas del área.
- Se espera fecha para la defensa de la tesis de doctorado.
- Los integrantes de esta línea de Investigación participan en el dictado de asignaturas/cursos de grado/postgrado en la Facultad de Informática de la UNLP y en otras universidades del país. En particular, en la UNLP, se dicta la asignatura “Calidad de Sistemas de Software”

5. BIBLIOGRAFIA

- [1] S. Esponda, A. Pasini, and P. Pesado, “Modelo de evaluación de datos utilizando el enfoque GQM,” vol. 25000, pp. 418–428.
- [2] ISO, “ISO/IEC 25000:2014 Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Guide to SQuaREtle,” 2014.
- [3] J. S. Preisegger, A. Greco, A. Pasini, M. Boracchia, and P. Pesado, “Marco de vinculación de datos abiertos aplicado al contexto de datos medioambientales,” pp. 684–694, [Online]. Available: <http://sedici.unlp.edu.ar/handle/10915/113243>.
- [4] E. Estevez, N. V. Lopes, and T. Janowski, “Smart Sustainable Cities - Reconnaissance Study. Operating Unit ON Policy-Driven. Electronic Governance,” *United Nations University, Canada*, 2017.
- [5] S. Esponda *et al.*, “Fortalecimiento de la calidad en procesos de software y Gobernanza Digital,” *Sedici.Unlp.Edu.Ar*, pp. 574–578, [Online]. Available: <http://sedici.unlp.edu.ar/handle/10915/77106>.
- [6] CEPAL Naciones Unidas. “Desde el gobierno abierto al Estado abierto en América Latina y el Caribe”2017
- [7] Millat, A. E. (2020). Ciudades Resilientes Formando un Mundo Digital.
- [8] R. Muñoz, A. Pasini, and P. Pesado, “Innovación en el Sector Público para Ciudades Inteligentes Sostenibles.”
- [9] E. Estevez *et al.*, “Desarrollo de Recursos Humanos para la Gobernanza de Ciudades Inteligentes Sostenibles – Proyecto CAP4CITY,” 2020.
- [10] A. Pasini, R. Muñoz, P. Pesado, and I. De Investigación, “Gestor de contenidos orientado a portales de organismos gubernamentales,” pp. 0–3.

Resolución de conflictos para el modelado estático orientado a aspectos

Laura Zeligueta, Marcelo Palma, Fernando Pincioli

Instituto de Investigaciones
Facultad de Informática y Diseño
Universidad Champagnat
Mendoza, Argentina, +54 (261) 424-8443
zeligueta@uch.edu.ar; mpalma@uch.edu.ar; fpincioli@aconcaguasf.com

Resumen

En este proyecto se está trabajando en el establecimiento de los lineamientos para la fase de requisitos de software del ciclo de vida del desarrollo de software (SDLC), poniendo especial énfasis en la vista estática y con el empleo del paradigma de orientación a aspectos, de manera de obtener los beneficios que ofrece [1].

Es el quinto de la misma línea de investigación de desarrollo de software orientado a aspectos (AOSD) que se vienen desarrollando en forma consecutiva desde el año 2013 en el Instituto de Investigaciones de la Facultad de Informática y Diseño de la Universidad Champagnat y que continúan de un proyecto que fuera desarrollado en la Facultad Regional Mendoza de la Universidad Tecnológica Nacional.

Palabras clave: orientación a aspectos, aspectos tempranos, requisitos de usuario, requisitos funcionales, separación de incumbencias, encapsulamiento de incumbencias, composición de incumbencias, resolución de conflictos, AOP4ST.

Contexto

Dentro de las líneas de investigación del Instituto de Investigaciones de la Facultad de Informática y Diseño de la Universidad Champagnat, se encuentra la línea de “Ingeniería de Software”. En ella, se llevaron a cabo los siguientes proyectos dedicados al

desarrollo de software orientado a aspectos (AOSD):

- “*Procesos de desarrollo de software de calidad basados en aspectos*”, realizado en conjunto con la UTN Facultad Regional Mendoza durante 2011 y 2012 [2].
- “*Definición de criterios para la detección temprana de aspectos en el modelado de negocios y el desarrollo de los requisitos*” llevado a cabo por completo en la Universidad Champagnat, desde 2013 a 2016 y presentado en WICC 2015 [3].
- “*Modelado de procesos de negocio orientados a aspectos con BPMN*”, de la Universidad Champagnat y realizado en los años 2016 a 2018, y que fue presentado en WICC 2016 [4] y 2017 [5].
- “*Ingeniería de requisitos orientada a aspectos en AOP4ST*”: este proyecto abordó la fase de requisitos de usuario y se desarrolló entre 2018 y 2020 [6].

El proyecto actual, que se denomina “Ingeniería de requisitos de software orientada a aspectos en AOP4ST”, se inició en julio de 2020 y ya se produjeron algunas publicaciones: [1] [7] [8]. Cuenta con el financiamiento de la Universidad Champagnat y de la empresa Aconcagua Software Factory S.A., una de las principales fábricas de software de la Argentina.

Introducción

AOP4ST, sigla derivada de Aspect-Oriented Process for a Smooth Transition, es un proceso marco para el SDLC. En su versión actual, está enfocado en las fases tempranas,

early aspects [9], compuesta por diferentes modelos: de negocio, de requisitos de usuario y de requisitos de software. Este último considera tres vistas: funcional, estática y de estados.

El acrónimo AOP4ST corresponde Aspect-Oriented Process for a Smooth Transition, que es un proceso marco para el desarrollo de software que pretende [8]:

- a) Ofrecer un proceso marco, no específico y liviano, de modo que permita su empleo con diferentes modelos del ciclo de vida del desarrollo del software (SDLC).
- b) Cubrir las etapas tempranas del SDLC, desde el modelado de negocio hasta la obtención de una especificación de requisitos completa y coherente.
- c) Emplear herramientas y técnicas estándares, de amplia difusión en la industria, para facilitar su adopción inmediata.
- d) Emplear notaciones estándares, para lograr modelos y especificaciones comprensibles y no ambiguas, que puedan contar con soporte de herramientas de software disponibles en el mercado.
- e) Desarrollar los productos intermedios de las etapas tempranas del SDLC empleando el paradigma de la orientación a aspectos.
- f) Obtener las incumbencias en forma progresiva a lo largo de todos los modelos.
- g) Mantener la separación de incumbencias a lo largo de todos los modelos.
- h) Mantener la trazabilidad bidireccional de las incumbencias de punta a punta.
- i) Obtener las incumbencias en forma natural a lo largo de todos los modelos, de manera de no afectar la obtención de los objetivos de cada modelo.

El presente proyecto de investigación pretende establecer lineamientos para la vista estática del modelo de requisitos de software a partir de la vista funcional realizada en trabajos anteriores, dejando la vista de estados como próximo objeto de estudio. Procura presentar una visión interna del dominio del problema al descomponerlo en las clases específicas que lo

constituyen, a diferencia del modelo de casos de uso, que presenta una visión externa del sistema [7].

Considerando las actividades propias del enfoque orientado a aspectos, mantiene la separación de incumbencias considerando los mismos paquetes que existen en la vista funcional del modelo de requisitos y elaborando dentro de ellos los diagramas de clases que materializan la vista estática.

En cuanto a la composición aplica un procedimiento sintáctico que combina las clases que inicialmente poseen el mismo nombre en las diferentes incumbencias. El total de los atributos, operaciones y relaciones es el resultado de la unión de todos estos elementos desde todas las incumbencias individuales y que terminan apareciendo por completo, y solo una vez, en la composición resultante. El proceso presenta distintas estrategias según se trate de incumbencias pares e impares [8].

En cuanto a la resolución de conflictos establece algunos acuerdos y consideraciones previas y preventivas tales como criterios de diagramación y nomenclatura.

Finalmente, propone buenas prácticas de manera de asegurar los objetivos de la fase del ciclo de vida en cuestión, la calidad del producto software en desarrollo y la obtención de los beneficios que son la razón del empleo del paradigma de orientación a aspectos.

Estos beneficios apuntan al desarrollo de un producto de software final, como así también de los productos intermedios que permiten producirlo, más modular, mantenible, reusable, extensible, comprensible, etc. [8], al administrar en forma separada las incumbencias que están desparramadas y enredadas en cada uno de los niveles de abstracción a lo largo del ciclo de vida completo.

Desde ya que los principales aportes a los requisitos tempranos (early aspects) con el

empleo de modelado estático, obtenidas mediante nuestro estudio de mapeo sistemático [1] y otros estudios de mapeo complementarios [10] [11] [12] [13], son fuentes obligadas de referencia y de las que nos nutrimos en este proyecto.

Líneas de Investigación, Desarrollo e Innovación

Los proyectos de investigación que se vienen desarrollando en nuestra Facultad presentan cuatro ejes de investigación, donde los tres primeros se ven integrados con el cuarto:

1. Modelado de procesos de negocio orientados a aspectos.
2. Separación y composición de incumbencias con resolución de conflictos.
3. Especificación de requisitos y gestión de incumbencias con casos de uso e historias de usuario.
4. Procesos de desarrollo de software orientados a aspectos.

Resultados y Objetivos

Se consideraron resultados obtenidos en proyectos anteriores como insumos para el trabajo actual, como así también la línea de trabajo que sigue AOP4ST. Tras haber detectado las primeras incumbencias pares e impares en el modelo de negocio [4] [5], con el empleo del enfoque asimétrico tradicional de la orientación a aspectos [14] [15] y el enfoque simétrico que se propuso más tarde [16], se sigue el desarrollo y la administración de los requisitos de usuario, que permiten enriquecer el modelo con nuevas incumbencias [17]. Estos requisitos de usuario son especificados mediante elementos vistas: funcional, estática y de estados [3].

La vista estática constituye el objeto de estudio de nuestro proyecto de investigación actual. Tiene como objetivo el profundizar los aportes que ella debe realizar a AOP4ST desde su rol en el proceso marco completo.

Hasta el momento, estos son los resultados alcanzados:

Detección y separación de incumbencias: desde el modelo de negocio y hasta el modelo de requisitos de software en su vista funcional fue posible mantener separadas las incumbencias. Pero al entrar en la vista estática, se observa que la mayoría de las clases involucradas en las funcionalidades (incumbencias) que se mantuvieron separadas hasta aquí, están presentes en más de una funcionalidad. Dicho de otra manera, los elementos de las vistas funcional y estática se relacionan de manera “muchos a muchos”, con lo que aparece nuevamente el enredo. Así, para poder seguir manteniendo la separación de las incumbencias en la vista estática, se consideró el concepto de “módulo de caso de uso” [18], que contiene un diagrama de clases para cada incumbencia en forma separada y que, por supuesto, luego se deberán componer. En nuestro proyecto trabajaron por separado cuatro analistas de sistemas elaborando diagramas de clases correspondientes a procesos que pertenecen a un laboratorio bioquímico.

Composición de incumbencias: se realizó teniendo en cuenta el proceso que define AOP4ST [8] obteniendo un único modelo sobre los diagramas de clases desarrollados por los diferentes analistas. Dado que hay diferentes formas de realizar la composición del modelo de clases, se siguieron las alternativas de mezcla y sobreescritura propuestas por Clarke y Baniassad [15], a las que incorporamos algunas variaciones y a las que sumamos una tercera solución, de selección, que comprobamos que ofrece resultados más alentadores. Durante este proceso es posible la aparición de nuevas clases que podrían suscitar nuevas incumbencias. Esto obliga a que las vistas estática y funcional se elaboren en forma iterativa e incremental.

Resolución de conflictos: Finalizada la composición de los diagramas se identificaron, analizaron y clasificaron los conflictos. En primer lugar, se elaboró un conjunto de acuerdos y recomendaciones para reducir el número de conflictos. Luego, los modelos se

reconstruyeron con el uso del conjunto de acuerdos de modelado, y se volvieron a componer, con una importante mejora en el número de conflictos detectados después de la segunda composición.

Creemos que hay mucho que hacer sobre las actividades pre y post composición[19]. Pensamos profundizar en el estudio sobre el impacto de los cambios que hacemos durante la composición: adición, eliminación, modificación y derivación de elementos en el modelo compuesto[20]. La producción de patrones de modelado también será muy útil. Las reglas y convenciones para nombrar elementos son cruciales ya que los nombres de los elementos de modelado son la base de la composición. El estudio de los diferentes tipos de conflictos y su clasificación también será una gran contribución al diseño de técnicas que permitan abordarlos y resolverlos de manera más efectiva. La composición de las relaciones es también un área importante para explorar[21]. Finalmente, la automatización de nuestras reglas colaborará con una composición eficiente de modelos.

Formación de Recursos Humanos

El proyecto promueve la formación del equipo de profesores, alumnos y egresados de la Universidad Champagnat.

Además, se están elaborando diferentes tesis sobre el cuerpo de esta investigación:

1. Fernando Pincirolí, Doctorado en Ciencias Informáticas, Universidad Nacional de San Juan. Tesis defendida, con la que obtuvo el grado de Doctor en Ciencias de la Informática en noviembre de 2020.
2. Marcelo Palma, Maestría en Negocios y Tecnología, Universidad Champagnat. Tesis en desarrollo.
3. Marcelo Fransoy, Maestría en Ingeniería en Sistemas de Información, Universidad Tecnológica Nacional, Facultad Regional Buenos Aires. Tesis en desarrollo.
4. Jerónimo Vargas, Maestría en Negocios y Tecnología, Universidad Champagnat. Tesis en desarrollo.
5. Pablo Ingrassia, Maestría en Negocios y Tecnología, Universidad Champagnat. Tesis en desarrollo.
6. Gustavo Albino, Maestría en Ingeniería de Software, Universidad Nacional de San Luis. Tesis en desarrollo.
7. Marcelo Fransoy, Especialización en Ingeniería en Sistemas de Información, Universidad Tecnológica Nacional, Facultad Regional Buenos Aires. Tesis en desarrollo.

Referencias

- [1] F. Pincirolí, J. L. Barros-Justo, and R. Forradellas, "Systematic Mapping Study: on the coverage of aspect-oriented methodologies for the early phases of the software development life cycle," *J. King Saud Univ. - Comput. Inf. Sci.*, 2020.
- [2] F. Pincirolí, "Consideraciones para un proceso de desarrollo de software de calidad orientado a aspectos," in *Sexto Encuentro de Investigadores y Docentes de Ingeniería EnIDI 2011*, 2011.
- [3] F. Pincirolí, "AOP4ST – Aspect-Oriented Process for a Smooth Transition," in *WICC 2015 - XVII Workshop de Investigadores en Ciencias de la Computación*, 2015.
- [4] F. Pincirolí and L. Zeligueta, "El modelo de negocio en AOP4ST," in *WICC 2016 - XVIII Workshop de Investigadores en Ciencias de la Computación*, 2016.
- [5] F. Pincirolí and L. Zeligueta, "Modelado de negocios orientado a aspectos con AOP4ST," in *WICC 2017 - XIX Workshop de Investigadores en Ciencias de la Computación*, 2017.
- [6] F. Pincirolí and M. Palma, "Desarrollo de requisitos en 'Aspect-Oriented Process for a Smooth Transition,'" in *XXII Workshop de Investigadores en Ciencias de la Computación - WICC 2020*, 2020.
- [7] F. Pincirolí, "Modeling the Static View in Aspect-Oriented Software Development," no. section 6.
- [8] F. Pincirolí, Proceso marco orientado a aspectos en las etapas tempranas del ciclo de vida del desarrollo de software para una transición en la industria, Universidad Nacional de San Juan, San Juan, 2020.
- [9] A. Rashid, P. Sawyer, A. Moreira, and J. Araújo, "Early aspects: A model for aspect-oriented requirements engineering," *Proc.*

- IEEE Int. Conf. Requir. Eng.*, vol. 2002-Janua, pp. 199–202, 2002.
- [10] A. Magableh, Z. Shukur, and N. M. Ali, “Systematic review on aspect-oriented UML modeling: A complete aspectual UML modeling framework,” *J. Appl. Sci.*, vol. 13, no. 1, pp. 1–13, 2013.
- [11] A. A. Magableh, “Systematic review of aspect-oriented formal method,” *Int. J. Comput. Appl. Technol.*, vol. 56, no. 2, pp. 132–140, 2017.
- [12] M. Guessi, L. B. R. Oliveira, and E. Y. Nakagawa, “Modeling Aspect-oriented Software Systems Using UML: A Systematic Mapping,” no. November 2016, 2010.
- [13] M. Guessi, L. B. Ruas Oliveira, and E. Yumi Nakagawa, “Extensions of UML to Model Aspect-oriented Software Systems,” *CLEI Electron. J.*, vol. 14, no. 1, pp. 1–18, 2011.
- [14] M. Katara and S. Katz, “A concern architecture view for aspect-oriented software design,” *Softw. Syst. Model.*, vol. 6, no. 3, pp. 247–265, Aug. 2007.
- [15] S. Clarke and E. Baniassad, *Aspect-oriented analysis and design. The Theme approach*. Boston: Addison-Wesley, 2005.
- [16] W. Harrison, H. Ossher, and P. Tarr, “Asymmetrically vs. Symmetrically Organized Paradigms for Software Composition,” vol. 22685, 2002.
- [17] F. Pincioli, “Requisitos de usuario y gestión de la demanda en AOP4ST,” in *XX Workshop de Investigadores en Ciencias de la Computación - WICC 2018*, 2018.
- [18] I. Jacobson and P. Ng, *Aspect-oriented software development with use cases*. Addison-Wesley, 2005.
- [19] R. Reddy, R. France, F. Fleury, and B. Baudry, “Model composition - a signature based approach,” *Proc. Asp. Oriented Model. Work. held with Model. 2005*, p. 2, 2005.
- [20] K. S. Farias de Oliveira, “Empirical Evaluation of Effort on Composing Design Models,” Pontificia Universidade Católica do Rio de Janeiro, 2012.
- [21] S. A. H. Elasri, E. Elabbassi, Semantic integration of UML class diagram with semantic validation on segments of mappings, 2018.

Medición para la evaluación transversal de metodologías de gestión de proyectos

Patricia Cristaldo, Lucas La Pietra, Soledad Retamar, Anabella De Battista

Grupo de Investigación en Bases de Datos, Departamento Ingeniería en Sistemas de Información,
Fac. Reg. Concepción del Uruguay, Universidad Tecnológica Nacional
Entre Ríos, Argentina
{cristaldop, lapietral, retamars, debattistaa}@frcu.utn.edu.ar

María Daniela López de Luise

CI2S Labs- Computational Intelligence & Information Systems Lab, Buenos Aires
daniela_ldl@ieee.org

Resumen

La gestión de proyectos de software incluye varios aspectos: dirección, alcance, interesados, riesgos, planificación y control de actividades, requerimientos del proyecto, y objetivos de negocios. Remite a las habilidades del director de proyecto para administrar problemas vinculados con la gestión y tecnología. La gestión correcta de proyectos busca la conclusión en tiempo y forma de los mismos. Entre las razones para no alcanzar los objetivos en tiempo y forma se pueden citar: planificación insuficiente, pobre definición de requerimientos, falta de habilidades, problemas con la disciplina de gestión y organización por parte de los encargados de llevar adelante los proyectos. Asimismo, diversos autores proponen el enfoque “híbrido”, que fusiona las propuestas tradicionales con ágiles. En esta metodología los gestores del proyecto deben centrarse no sólo en el objetivo final sino también en el momento donde se opta por una metodología u otra.

En este artículo se presentan los tópicos de interés del proyecto *Generación de marco de medición para la evaluación transversal de metodologías de gestión de proyectos*.

Palabras clave: Gestión de proyectos de software, métricas de gestión, minería de datos, procesamiento de lenguaje natural.

Contexto

El presente trabajo se desarrolla en el ámbito del proyecto *Generación de marco de medición para la evaluación transversal de metodologías de gestión de proyectos (SIECACU0008388)* del Grupo de Investigación en Bases de Datos, perteneciente al Departamento Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional, Facultad Regional Concepción del Uruguay.

1. Introducción

La gestión de proyectos de software, comprende la fusión de la Ciencia y la Gestión o Management. Ambos términos, gestión y proyecto, se desarrollan como un conjunto de actividades esencialmente útiles, funcionales y alineadas, con énfasis en la objetividad y con un enfoque en las técnicas y el control. Por ello, la gestión de proyectos es un esfuerzo coordinado para alcanzar una meta en particular o realizar una función específica [1]. Incluye varios aspectos: dirección, alcance, interesados, riesgos, planificación y control de actividades, requerimientos del proyecto, y objetivos de negocios. Dependiendo del tipo de proyecto y de la organización, la gestión es llevada a cabo por un gerente, administrador, director o jefe de proyecto, quien es la persona que se encarga de lograr los objetivos del proyecto. Además de las habilidades específicas en el

área que está inmerso el proyecto, debe poseer conocimientos y competencias para gestionar proyectos, y actitudes básicas de personalidad y liderazgo.

Como ayuda, existen en el mercado numerosas metodologías y guías de gestión de proyectos. Algunas de ellas son PMBOK [1], PRINCE2 [2][3], APM [4], ISO 21500 [5], SCRUM [6] [7], KANBAN [8], CRISP-DM [9] [10]. La gestión correcta de proyectos busca la conclusión en tiempo y forma de los mismos [11]. Según el informe CHAOS del año 2020, el 28% de los proyectos respetan el tiempo, presupuesto, características y funciones requeridas. En contraste, el 32% no respeta alguno de los ejes. El 53% de los proyectos sufren retrasos, exceden el presupuesto, o implementan menos requerimientos [12]. Esto supera un 15% lo informado en año 2010. Asimismo, el informe muestra una cancelación del proyecto sin producto del 21%.

Entre las razones para no alcanzar los objetivos en tiempo y forma se pueden citar: planificación insuficiente [13] [14] [15] pobre definición de requerimientos [16] [17] [18], falta de habilidades, problemas con la disciplina de gestión y organización por parte de los encargados de llevar adelante los proyectos [19].

Diversos autores proponen el enfoque “híbrido”, que fusiona las propuestas tradicionales con ágiles [20] [21] [22] [23][24]. En esta metodología los gestores del proyecto deben centrarse no sólo en el objetivo final sino también en el momento donde se opta por una metodología u otra.

Estos cambios en las gestiones, han sido acompañados de numerosos esfuerzos para evaluar de manera sistemática la calidad de las gestiones y sus efectos. Algunos de dichos esfuerzos se presentan y analizan en este artículo, junto con las propuestas para métricas e indicadores con el objetivo de medir y evaluar todos los aspectos y componentes de las metodologías y guías de gestión de proyectos. Se pretende así desarrollar un marco de medición, a través de las métricas consignadas, para evaluar transversalmente diferentes metodologías

cuantitativamente. Esto permite, entre otras cosas, expresar el grado de aplicabilidad de las estrategias en las distintas fases de un proyecto y/o en proyectos de diferentes contextos.

Este proyecto se encuentra alineado al desarrollo de métricas, pero no para la gestión de proyectos en sí, sino para la evaluación del proceso de aplicación de modelos de gestión. Asimismo, al buscar una métrica integral de gestión de proyectos, se aportan nuevas métricas de carácter original, basadas en procesamiento de lenguaje natural, de forma tal traducir los requerimientos del texto, necesarios para la gestión de los proyectos [25][26] [27].

2. Líneas de Investigación, Desarrollo e Innovación

La línea de trabajo principal de nuestro proyecto de investigación es definir y desarrollar un marco de medición para la determinación de aplicabilidad y evaluación transversal de las metodologías de gestión de proyectos. Conteniendo métricas e indicadores vinculados a las áreas de conocimiento de la guía para la gestión de proyectos PMBOK [1]: gestión del alcance, gestión del tiempo, gestión de costos, gestión de la calidad, gestión de recursos, gestión de interesados, gestión de las comunicaciones, gestión de riesgos y gestión de las adquisiciones.

Según la guía de buenas prácticas en gestión de proyectos, la gestión del alcance es el conjunto de procesos necesarios para asegurar que se incluya todo el trabajo requerido para completar el proyecto satisfactoriamente [1]. Teniendo esto en cuenta, es importante formalizar el documento inicial del proyecto, *Enunciado del Alcance del Negocio* con una descripción completa de la lista de requerimientos, de forma tal de sentar las bases sólidas de un proyecto de calidad [1]. De acuerdo a la literatura presentada en la sección anterior, es importante determinar el nivel de comprensibilidad en el Enunciado del Negocio (misión, visión, alcance). Esto implica:

- Generar una descripción correcta del alcance de un proyecto, es decir, la misma debe contener una cantidad sustantivos y verbos igual a los contenidos en el alcance del proyecto, y que se repiten en la totalidad de los requisitos.
- Calcular la completitud en la descripción del requisito, contenida explícitamente.

Asimismo del alcance del proyecto se deben extraer los distintos tipos de:

- *Requerimientos del Negocio:* conjunto de necesidades y oportunidades de la organización
- *Requerimientos de los Interesados:* conjunto de necesidades de quienes participan en el proyecto.
- *Requerimientos de Soluciones:* constituyen el conjunto de características y funcionalidades de producto o servicio.
- *Requerimientos del Proyecto:* conjunto de acciones y procesos que el proyecto debe proveer.
- *Requerimientos de Calidad (RQC):* conjunto de condiciones o criterios que el producto debe satisfacer.

Otro aspecto importante es calcular la graduación de consistente. La descripción de requisitos contradictoria, usa partículas negativas del lenguaje ante otra positiva para el mismo sustantivo o verbo. Además, la descripción de requisitos debe permitir inspección, análisis, demostración y/o pruebas. Por otro lado, se debe poder determinar la trazabilidad de los requisitos, esto es, la descripción de requisitos contener estados (aprobado, diseñado, implementado, testeado, entre otros).

Se pretende desarrollar un marco de medición, a través de las métricas consignadas, para evaluar transversalmente diferentes metodologías cuantitativamente. Esto permite, entre otras cosas, expresar el grado de aplicabilidad de las estrategias en las distintas fases de un proyecto y/o en

proyectos de diferentes contextos. Estas métricas se corresponden a un conjunto de métricas consignadas para evaluar metodologías de gestión de proyectos, en [28][29][30] se han publicado otras que preceden.

3. Resultados obtenidos y esperados

Con este proyecto se espera definir y desarrollar un marco de medición para la determinación de aplicabilidad y evaluación transversal de metodologías de gestión de proyectos, a partir de expresiones textuales de la descripción de un proyecto. Además se espera obtener una herramienta eficiente para la captura de la información de las encuestas a realizar. Asimismo se pretende relacionar las métricas para la identificación del conocimiento implícito que se encuentra dentro de los textos que describen el alcance del proyecto, aplicando técnicas de minería de textos, haciendo referencia al proceso de descubrir y extraer conocimiento relevante y no trivial a partir de textos no estructurados. Esto es notable dado que la complejidad del lenguaje natural dificulta el acceso a la información en los textos y aún se está lejos de poder construir representaciones de significado de propósito general a partir de texto sin restricciones.

4. Formación de Recursos Humanos

Este proyecto dio inicio a una nueva línea de investigación dentro del Grupo de investigación en Bases de Datos de la Fac. Reg. Concepción del Uruguay de la U.T.N. Dos de los investigadores del proyecto están desarrollando tesis doctorales y uno, tesis de maestría. En el proyecto colaboran dos becarios graduados con beca de iniciación a la investigación, que tienen previsto la realización de posgrados en el área temática del proyecto. Además participan en el proyecto tres becarios alumnos de la carrera Ingeniería en Sistemas de Información que inician su formación en la investigación. Se prevé la incorporación de alumnos de grado

para que desarrollen su trabajo de práctica supervisada.

5. Referencias

- [1] Project Management Institute. A Guide to the Project Management Body of Knowledge. 6 Ed. ISBN- 10: 9781628251845, 2017.
- [2] PRINCE2. An introduction to PRINCE2: managing and directing successful projects. Office of Government Commerce. Stationery Office, 123 p. ISBN-10: 0113311885, ISBN-13: 978-0113311880.
- [3] Böhm, A. Application of PRINCE2 and the Impact on Project Management. ISBN (eBook) 978-3-640-42634-8, 2009.
- [4] Highsmith, J. Agile project management: creating innovative products. 2 Ed. Boston, MA: Addison-Wesley. 432 p. 2010.
- [5] ISO. ISO 21500:2012 Guidance on Project Management, Geneva: ISO, 2012.
- [6] Sutherland, J. Scrum: The art of doing twice the work in half the time. New York: Crown Business. 256 p. ISBN-10: 038534645X, ISBN-13: 978-0385346450, 2014.
- [7] Van Solingen R., Van Lanen R. (Scrum for managers) Scrum voor Managers. Den Haag: Academic Service. EAN: 9789012585903, 2014.
- [8] Lei H., Ganjeizadeh F., Jayachandran P., Ozcan, P. A statistical analysis of the effects of Scrum and Kanban on software development projects. Robotics and Computer-Integrated Manufacturing, <http://dx.doi.org/10.1016/j.rcim.2015.12.001>, 2015.
- [9] Shearer C. The CRISP-DM model: The new blueprint for data mining. Journal of Data Warehousing, no. 4, pp. 13-22, 2000.
- [10] Shafique U., Qaiser H. *A Comparative Study of Process Models Data Mining (KDD, CRISP-DM and SEMMA)*. International Journal of Innovation and Scientific Research, pp.217-222, 2014.
- [11] Varajão, J. Dominguez, C. Ribeiro, P. Paiva, A. Critical success aspects in Project management: similarities and differences between the construction and software industry, Tech. Gazette 21 (3) 583–589, 2014.
- [12] TSG. The CHAOS Report. The Standish Group, 2018. Disponible en <https://secure.standishgroup.com/reports/flyers/CM2018-TOC.pdf>, 2018.
- [13] Lehtinen T., Mäntylä M., Vanhanen J., Itkonen J., Lassenius, C. Perceived causes of software project failures – An analysis of their relationships. Information and Software Technology, pp. 623–643, 2014.
- [14] Ramos P., Mota C. Perceptions of success and failure factors in information technology projects: a study from Brazilian companies. Procedia - Social and Behavioral Sciences, pp. 349 – 357, 2014.
- [15] Montequin S., Fernandez C., Fernandez O., Balsera J. *Analysis of the Success Factors and Failure Causes in Projects: Comparison of the Spanish Information y Communication Technology (ICT) Sector*. Journal Information Technology Project Management, pp. 18-31, 2016.
- [16] Chow T., Chao D. A survey of critical success factors in agile software projects. Journal of systems and software, Jun. Available: Science Direct, pp.961–971, 2008.
- [17] Elkadi H. Success and failure factors for e-government projects: A case from Egypt. Egyptian Informatics Journal, pp. 165–173, 2013.

- [18] ElEmam K., Koru A. A replicated survey of IT software project failures Software. IEEE Software, pp. 84–90. 2008.
- [19] Blaskovics B. The impact of project manager on project success – The case of ICT sector. Assistant Professor, Strategy and Project Management Department, Corvinus University Budapest E-mail: balint.blaskovics@uni-corvinus.hu, 2016.
- [20] Esteki M., Gandomani T., Farsani H. A Risk Management Framework for Distributed Scrum Using PRINCE2 Methodology. Bull. Electrical Engineering and Informatics. DOI: 10.11591/eei.v9i3.1905. 2020.
- [21] Jabar M., Mohd Ali N., Jusoh Y., Abdullah S., Mohanarajah S. A pilot examination of an improved agile hybrid model in managing software projects success. *Test Engineering and Management*, pp. 3040-3046, 2019.
- [22] Smoczyńska A., Pawlak M., Poniszewska-Marańda, A. Hybrid agile method for management of software creation. *Advances in Intelligent Systems and Computing*, pp. 101-115. 2019.
- [23] Włodzimierz Wysockia W., Orłowski, C. A multi-agent model for planning hybrid software processes. *Procedia Computer Science* pp. 1688–1697, 2019.
- [24] Mousaei M., Gandomani T. A new project risk management model based on Scrum framework and Prince2 methodology. *International Journal*, 2018.
- [25] Martino Di Giuda G., Locatelli M., Schievano M., Pellegrini L, Pattini G., Ettore Giana P., Seghezzi E. Natural Language Processing for Information and Project Management. *Digital Transformation of the Design, Construction and Management Processes of the Built Environment*, Research for Development, https://doi.org/10.1007/978-3-030-33570-0_9, 2020.
- [26] Partha Sarathy B., Abhijay G., Aditya G., Baisakhi C. Natural Language Information Extraction Through Non-Factoid Question and Answering System (NLIEQA Non-Factoid). DOI: https://doi.org/10.1007/978-981-15-3383-9_10, 2020.
- [27] Hassan F., Le T. Automated Requirements Identification from Construction Contract Documents Using Natural Language Processing, DOI: 10.1061/(ASCE)LA.1943-4170.0000379, 2020.
- [28] Cristaldo P, López De Luise D, La Pietra L, De Battista A, Hemanth J. Metrics for the Systematic Evaluation of Project Management Methodologies. *International Journal of Service Science, Management, Engineering, and Technology (IJSSMET)*, 2021.
- [29] Cristaldo P, López De Luise D, La Pietra L, De Battista A. Metrics for validation and traceability of Project Management Requirements. *Global Research and Development Journal for Engineering (GRDJE) ISSN (online): 2455-5703*, 2021.
- [30] Cristaldo P, López De Luise D, La Pietra L, De Battista A. Adjustment Level Metrics for Requirements in Project Management. *Proceedings of 2021 8th Intl. Conference on Soft Computing & Machine Intelligence (ISCMi 2021)*, IEEE ISSN: 2640-0146, ISBN: 978-1-7281-8682-5, IEEE Catalog Number: CFP2144Z-USB, Cairo, Egypt.

ASPECTOS DE INGENIERÍA DE SOFTWARE, BASES DE DATOS RELACIONALES, NO RELACIONALES Y COMO SERVICIOS EN LA NUBE PARA EL DESARROLLO DE SISTEMAS DE SOFTWARE HÍBRIDOS.

Luciano Marrero^{ID}, Pablo Thomas^{ID}, Ariel Pasini^{ID}, Rodolfo Bertone^{ID}, Eduardo Ibáñez^{ID}, Verónica Aguirre^{ID}, Marisa Panizzi¹, Verena Olsow^{ID}, Fernando Tesone^{ID}, Patricia Pesado^{ID}

Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata
50 y 120 La Plata Buenos Aires

Centro Asociado CIC

¹ Universidad de Morón

{lmarrero, pthomas, apasini, pbertone, eibanez, vaguirre, volsow, ftesone, ppesado}@lidi.info.unlp.edu.ar,

¹marisapanizzi@outlook.com

RESUMEN

Se presenta una línea de investigación que tiene por objeto estudiar las problemáticas actuales que afronta los procesos de diseño e implementación del Software (Ingeniería de Software y Bases de Datos) ante una gran variedad aplicaciones multiplataforma utilizadas por millones de usuario simultáneamente. En la actualidad la mayoría de las aplicaciones de Software son multiplataforma y requieren nuevas formas de pensar los diseños de Bases de Datos en complemento con las tradicionales metodologías ágiles. El modelo relacional de Bases de Datos (Codd 1970) [5], es el modelo predominante de almacenamiento de información. Sin embargo, la idea de considerar que un único modelo de datos pueda adaptarse de forma eficiente a todos los requerimientos, ha sido discutida. Surgen así, otros motores de Bases de Datos que poseen implementaciones propias no relacionales y se denominan Bases de

Datos NoSQL (No solo SQL) [8, 9, 13]. Estas Bases de Datos facilitan el almacenamiento de datos semiestructurados o sin estructura. Actualmente, existen una gran variedad de Sistemas de Gestión de Bases de Datos (SGBD) que implementan y almacenan sus datos de manera no estructurada. Estas implementaciones pueden ser para instalar de forma local o en la nube [10, 11].

La tecnología móvil impulsó aspectos que hace algunos años no eran considerados en el desarrollo del Software, tales como, movilidad, geolocalización, generación de grandes volúmenes de información y la diversidad de los dispositivos electrónicos involucrados [1, 2, 9, 12].

Palabras claves: Diseño no estructurado de datos, Metodologías ágiles, Bases de Datos Relacionales, Bases de Datos NoSQL, Bases de Datos como servicio en la nube, Aplicaciones Multiplataforma, Aplicaciones Móviles.

CONTEXTO

Esta línea de Investigación forma parte del proyecto (2018-2021) “Metodologías, técnicas y herramientas de ingeniería de software en escenarios híbridos”. Mejora de proceso.”, en particular del subproyecto “Ingeniería de Software para escenarios híbridos”, del Instituto de Investigación en Informática LIDI (III-LIDI) de la Facultad de Informática UNLP, acreditado por el Ministerio de Educación de la Nación.

Hay cooperación con Universidades de Argentina y se está trabajando con Universidades de Europa en proyectos financiados por el Ministerio de Ciencia y Tecnología de España y la AECID.

Se utilizan los recursos de Hardware y Software disponibles en el III-LIDI para diseñar, desarrollar y probar diferentes soluciones a problemáticas relacionadas con escenarios a investigar. Como resultado de esto, se espera obtener métricas reales que sirvan como referencia para los investigadores en la comparación de resultados.

Las publicaciones científicas generadas y la transferencia continua de resultados concretos, validan esta línea de investigación.

1. INTRODUCCIÓN

“Metodologías, técnicas y herramientas de Ingeniería de Software en escenarios híbridos. Mejora de proceso” del Programa de Incentivos, es un proyecto que propone profundizar las investigaciones que se vienen realizando en el III-LIDI y extender la mirada a nuevos desafíos y cambios que están en gestación. Se organiza en tres subproyectos que permiten atender de manera ordenada el objetivo general propuesto:

“SP1 - Ingeniería de Software para escenarios híbridos”, SP2 - Gobernanza Digital. Mejora de Procesos.” Y “SP3 - Metodologías y herramientas para la apropiación de tecnologías digitales en escenarios educativos híbridos [19].

Este artículo se centra en el subproyecto “SP1 - Ingeniería de Software para escenarios híbridos”. Se orienta a la investigación de metodologías y técnicas de la Ingeniería de Software, Bases de Datos Relacionales, Bases de Datos NoSQL y Bases de Datos como servicio en la nube, diseño de datos no estructurados, aplicaciones multiplataforma, escenarios híbridos, aplicaciones móviles, entre otras temáticas [20].

Con el crecimiento en el alcance y uso de internet y de los dispositivos móviles, sumado a la aparición de las redes sociales, se genera un crecimiento exponencial en el volumen de datos que hay que administrar y además, con una gran variedad de formatos y estructuras posibles. Se presentan así, un conjunto de nuevas alternativas y desafíos para la Ingeniería de Software, que van desde el diseño hasta la puesta en producción del Software. Los Sistemas Gestores de Bases de Datos (DBMSs su sigla en inglés) NoSQL representan una alternativa en la evolución del almacenamiento de datos, complementándose con una generación de tecnologías móviles y Web que debe responder eficientemente a las exigencias del usuario [6, 8, 11].

La exigencia del usuario actual pretende que las aplicaciones respondan de manera óptima todo el tiempo (rápida y eficazmente) y desde cualquier punto geográfico en cual se tenga acceso a Internet. Con el objetivo de satisfacer tales pretensiones, las empresas que desarrollan aplicaciones que son mundialmente utilizadas,

han tenido que poner foco en la escalabilidad de sus recursos enfrentando nuevos retos para el desarrollo de sus aplicaciones y para los proveedores de Bases de Datos [7, 14].

Las tecnologías actuales de Bases de Datos se han enfocado en las necesidades del usuario final generando nuevas alternativas, Bases de Datos No Relacionales (NoSQL) y las Bases de Datos como servicios en la Nube. Las Bases de Datos tradicionales (Relacionales), también poseen alternativas importantes para la tecnología móvil [10, 11, 12]. Existen distintos Sistemas de Bases de Datos Relacionales para el Software móvil, entre ellos, podemos encontrar a:

SQLite, es una librería que implementa un motor de Base de Datos autocontenido (embebido). Tiene Licencia Public Domain, y puede ser utilizada tanto en el desarrollo nativo de aplicaciones (Android o iOS), como también en aplicaciones multiplataformas (híbrido, interpretado o compilación cruzada) [9, 14].

Interbase, es embebido, con licencia comercial y conforme al estándar SQL. Se puede utilizar en el desarrollo nativo en Android e iOS [9].

SAP SQL Anywhere, integra un paquete de DBMSs relacionales y tecnologías de sincronización para servidores, en entornos de escritorio y móviles. Se puede utilizar en el desarrollo nativo, tanto en Android como en iOS [9, 10, 19].

SQLBase, desarrollado por la empresa Opentext. Tiene licencia de uso comercial. Se encuentra disponible para el desarrollo de aplicaciones móviles nativas, en Android e iOS [9, 14, 15].

En el caso de los motores de Bases de Datos NoSQL, también poseen sus variantes para la tecnología móvil, entre ellas:

Couchbase Lite (documental), es un DBMS embebido, utiliza JSON como formato de los documentos. Tiene licencia dual, y es posible utilizarlo para el desarrollo de aplicaciones móviles nativas Android e iOS, y multiplataforma (enfoques híbrido, interpretado y compilación cruzada) [9, 16, 19].

Firebase Realtime Database (documental), alojado en la nube, almacena los datos en un único JSON, y cuenta con sincronización de datos en tiempo real, manteniéndose disponibles aún sin conexión. Se encuentra disponible para el desarrollo de aplicaciones nativas en Android e iOS, y en el desarrollo de aplicaciones móviles (híbridas, interpretadas, y generadas por compilación cruzada) [9, 15].

Google Cloud Firestore, es un DBMS alojado en la nube, almacena los datos en documentos JSON, cuenta con sincronización de datos en tiempo real y mantiene los datos disponibles aún sin conexión. Es posible utilizarlo en el desarrollo de aplicaciones móviles nativas Android e iOS, y en el desarrollo de aplicaciones móviles (híbridas, interpretadas y generadas por compilación cruzada).

Oracle Berkeley DB, es una familia de Bases de Datos de Clave-Valor embebidas. Tiene licencia open source y se encuentra disponible para el desarrollo de aplicaciones nativas Android e iOS [9, 7, 10].

Como parte de las investigaciones realizadas, el análisis del proceso de despliegue del Software es otro aspecto de estudio. Las PyMES en Argentina representan casi el 80% de la industria del software y dada la necesidad de ser competitivas deben mejorar sus métodos y procesos de trabajo. En la mayoría de las empresas el proceso de despliegue no se realiza de manera sistemática y controlada, esto impacta

en la finalización del proyecto y la no aceptación del producto final, ocasionando inconvenientes que generan rehacer el trabajo y baja de productividad en su proceso. Ante estas dificultades, se propone realizar un modelo de proceso de despliegue de sistemas de software que permita a las PyMES mejorar la ejecución del proceso de despliegue [4, 17, 18].

La comunicación y la sincronización del trabajo continúan siendo un pilar fundamental para el éxito de un proyecto. La utilización de repositorios de información, por ejemplo, GIT, permiten realizar un control de versiones distribuido, trabajando en modo offline o en modo online, con la facilidad de disponer herramientas específicas para la resolución de conflictos entre versiones [3, 4, 6 y 16].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

- Metodologías de diseño para Bases de Datos No Relacionales.
- Investigar distintos tipos de DBMSs para aplicaciones móviles.
- Metodologías y técnicas de la Ingeniería de Software y su aplicación en el desarrollo de software para escenarios híbridos.
- Desarrollo de casos de estudio y métricas de performance para distintos tipos de motores de Bases de Datos No Relacionales (NoSQL) como por ejemplo MongoDB, Apache Cassandra, Redis, Neo4j, entre otros.
- Desarrollo de casos de estudio y métricas de performance para Bases de Datos en la nube. Cloud Firestore (Google), MongoDB Atlas (MongoDB), DataStax Astra (Apache Cassandra), entre otras.
- Investigar, desarrollar casos de estudio e incorporar nuevas pruebas de performance de otros tipos de motores de Bases de Datos, como, por ejemplo: NewSQL y Bases de Datos de Serie Temporales.
- Repositorios GIT.
- Modelo de procesos para el despliegue / puesta en producción de sistemas de software

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados esperados/obtenidos se pueden resumir en:

- Capacitación continua de los miembros de las líneas de investigación.
- Estudio y análisis de metodologías de diseño para Bases de Datos No Relacionales.
- Estudio y análisis de distintos DBMS Bases de Datos para el desarrollo de aplicaciones móviles [9].
- Estudio y análisis sobre la sincronización de requerimientos no funcionales para una Base de Datos para aplicación móvil con una Base de Datos *backend* [9, 10, 11].
- Estudio y análisis de Bases de Datos en la nube y otros tipos de Bases de Datos como NewSQL, Series Temporales, entre otras [10, 11, 12].
- Análisis de métricas para diversos casos de estudio realizado entre distintos tipos de motores de Bases de Datos.

- Estudio y análisis sobre escalamiento horizontal para pruebas de performance entre distintos motores de Bases de Datos [8, 13].
- Definición de procesos de Gestión de Incidencias utilizando repositorios GIT [11, 12].
- Análisis de metodologías para la interoperabilidad de sistemas web y aplicaciones móviles [11, 12].

Algunas de las transferencias realizadas por el III-LIDI relacionadas con este proyecto, entre otras, son:

- Aplicación móvil para la comunidad de la Facultad de Informática de la UNLP con información sobre horarios de finales, planes de estudio, calendario académico, las clases en tiempo real y las últimas novedades.
- Diseño de Bases de Datos para Instituciones Provinciales.
- Diseño y Gestión de Sistemas de congresos, utilizado por la RedUNCI y por otras entidades.
- Sistemas de Gestión Administrativa de Instituciones Universitarias (Sistema de inscripción y seguimiento de alumnos de la Facultad de Informática, Sistema de gestión administrativa de asignaturas, Sistema de gestión de asistencia de alumnos mediante QR, entre otros).

4. FORMACIÓN DE RECURSOS HUMANOS

Los integrantes de esta línea de investigación dirigen Tesinas de Grado y Tesis de Postgrado en la Facultad de Informática, y Becarios III-

LIDI en temas relacionados con el proyecto. Además, participan en el dictado de asignaturas/cursos de grado y postgrado de la Facultad de Informática de la UNLP.

5. BIBLIOGRAFÍA

1. Ingeniería de Software. Un Enfoque Práctico. Séptima Edición. Roger S. Pressman. McGraw-Hill Interamericana Editores, S.A. 2010. ISBN: 978-607-15-0314-5
2. Ingeniería de Software Teoría y Práctica. Pfliegger Shari Lawrence. Pearson / Prentice Hall. 2002. ISBN: 9789879460719
3. Ingeniería de Software. Novena Edición. Ian Sommerville. Addison Wesley / Pearson. 2011. ISBN: 978-607-32-0603-7
4. Ingeniería de Software Clásica y Orientada a Objetos. Sexta Edición. Stephen R. Schach. Mc Graw Hill Interamericana Editores S.A. 2006. ISBN: 970-10-5636-1.
5. Diseño Conceptual de Bases de Datos, un enfoque de entidades-interrelaciones. Carlo Batini, Stefano Ceri, Shamkant B. Navathe.. Addison-Wesley / Díaz de Santos. ISBN 0-201-60120-6 (1994).
6. Administración de Proyectos. Guía para el aprendizaje. Francisco Rivera Martínez, Gisel Hernández Chávez. Prentice Hall / Pearson. 2010. ISBN: 978-607-442-620-5.
7. NoSQL: modelos de datos y sistemas de gestión de bases de datos. XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018). Universidad Nacional del Nordeste (UNNE). <http://sedici.unlp.edu.ar/handle/10915/6725>

8. Análisis de performance en Bases de Datos NoSQL y Bases de Datos Relacionales. Luciano Marrero, Verena Olsowy, Fernando Tesone, Pablo Thomas, Lisandro Delia y Patricia Pesado. XXVI Congreso Argentino de Ciencias de la Computación (CACIC 2020). Universidad Nacional de La Matanza, del 5 al 9 de octubre del 2020. <https://cacic2020.unlam.edu.ar/es-ar/>
9. Un Análisis Experimental de Sistemas de Gestión de Bases de Datos para Dispositivos Móviles. 2021. Tesone Fernando, Thomas Pablo, Marrero Luciano, Olsowy Verena, Pesado Patricia. XXVII Congreso Argentino de Ciencias de la Computación (CACIC) (Modalidad virtual, 4 al 8 de octubre de 2021). Universidad Nacional de Salta (UNSA), Argentina. <http://sedici.unlp.edu.ar/handle/10915/130353>
10. Aspectos de ingeniería de software, bases de datos relacionales y bases de datos no relacionales para el desarrollo de sistemas de software en escenarios híbridos. Luciano Marrero, Pablo Thomas, Ariel Pasini, Rodolfo Bertone, Eduardo Ibáñez, Verónica Aguirre, Verena Olsowy, Fernando Tesone, Patricia Pesado. XXII Workshop de Investigadores en Ciencias de la Computación (WICC 2020). Universidad Nacional de la Patagonia Austral (UNPA). El Calafate, Santa Cruz (Mayo 2020). <http://sedici.unlp.edu.ar/handle/10915/104026>.
11. Aspectos de ingeniería de software, bases de datos relacionales, y bases de datos no relacionales y bases de datos como servicios en la nube para el desarrollo de software híbrido. Marrero Luciano, Thomas Pablo, Olsowy Verena, Pasini Ariel, Bertone Rodolfo, Ibáñez Eduardo, Aguirre Verónica, Panizzi Marisa, Pesado Patricia. XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021, Chilecito, La Rioja).
12. Aspectos de ingeniería de software y bases de datos para el desarrollo de sistemas de software en escenarios híbridos. Patricia Pesado, Rodolfo Bertone, Pablo Thomas, Ariel Pasini, Luciano Marrero, Eduardo Ibáñez, Alejandra Rípodas, Verónica Aguirre, Verena Olsowy, Fernando Tesone. XXI Workshop de Investigadores en Ciencias de la Computación (WICC 2019). Universidad Nacional de San Juan (UNSJ). (Abril 2019). <http://sedici.unlp.edu.ar/handle/10915/77088>.
13. Un estudio comparativo de bases de datos relacionales y bases de datos NoSQL. Pesado Patricia, Thomas Pablo, Delia Lisandro, Marrero Luciano, Olsowy Verena, Tesone Fernando, Fernández Juan Sosa. XXV Congreso Argentino de Ciencias de la Computación (CACIC 2019). Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019. ISBN 978-987-688-377-1. <http://sedici.unlp.edu.ar/handle/10915/91403>.
14. A. Nori, "Mobile and embedded databases," in Proceedings of the 2007 ACM SIGMOD International Conference on Management of data, pp. 1175-1177, 2007.
15. Q. H. Mahmoud, S. Zanin, and T. Ngo, "Integrating mobile storage into database systems courses," in Proceedings of the 13th

- annual conference on Information technology education, pp. 165-170, 2012.
16. Aplicaciones para Dispositivos Móviles. Estrategias y enfoques de desarrollo. Thomas Pablo Javier, et. al. XXI Workshop de Investigadores en Ciencias de la Computación (San Juan 2019). ISBN: 978-987-3984-85-3
 17. Which Change Sets in Git Repositories Are Related?. Ramadani, J., Wagner, S. EEE 2016. International Conference on Software Quality, Reliability and Security (Viena, Austria.)
 18. Synchronization and replication in the context of mobile applications. STAGE, A. (2005 Joint Advanced Student School Course 6: Next-Generation User-Centered).
 19. S. Lee, "Creating and using databases for android applications," International Journal of Database Theory and Application, vol. 5, no. 2, 2012..
 20. III-LIDI:
<http://weblidi.info.unlp.edu.ar/wp/proyectos/investigacion/>

ANÁLISIS Y RESOLUCIÓN DE LOS PROBLEMAS ASOCIADOS AL DISEÑO DE SISTEMAS DE IOT

Sebastián U. Flores, Mario Berón, Daniel Riesco

Departamento de Informática - Facultad de Ciencias Físico Matemáticas y Naturales
Universidad Nacional de San Luis - Ejército de Los Andes 950 - San Luis - Argentina
s.flores@outlook.com.ar, { mberon, driesco }@unsl.edu.ar

RESUMEN

Al momento de diseñar un sistema de IoT, sin importar si se parte desde un sistema existente que trabaja de forma offline, o si se desea crear un sistema desde sus inicios, se presentarán los siguientes desafíos:

En primer lugar, los sistemas de IoT pueden estar conformados por una amplia variedad de dispositivos, cada uno utilizando diferentes protocolos de comunicación y medios físicos para el establecimiento de la misma. Además, los dispositivos podrían encontrarse en ubicaciones geográficas muy distantes, en las que estén regidos por diferentes sistemas legales, y en las cuales la estructura de costos asociada a la conectividad entre los mismos sea muy diferente.

Por otra parte, la selección del hardware asociado a cada dispositivo puede variar dependiendo de los riesgos asociados a la actividad en la que se los involucre; de los costos asociados a la adquisición, instalación y mantenimiento en la región geográfica donde se los despliegue; de los protocolos de comunicación que se deseen utilizar; del nivel de calidad deseada en el desempeño de cada dispositivo; y de otros factores técnicos o comerciales. La selección de las tecnologías de Software a utilizar en cada dispositivo podría depender de factores similares a aquellos mencionados en la selección del hardware.

Además de estudiar las necesidades particulares de cada dispositivo, debe analizarse la arquitectura general del sistema de IoT. Esta arquitectura debe contemplar las diferentes formas de conectar a los dispositivos entre sí; las jerarquías de dispositivos; los servidores Web involucrados; los proveedores de servicios que serán contratados; los medios de

almacenamiento, procesamiento y publicación de la información; las personas involucradas y los demás componentes internos o externos que interactúan en el sistema.

Todas las consideraciones mencionadas previamente deben realizarse dentro de un marco de trabajo que garantice la privacidad y seguridad de la información tratada. Es por ello que en algunas regiones geográficas se han establecido diferentes legislaciones asociadas al tema, las cuales deben ser consideradas desde el comienzo del diseño del sistema de IoT. No obstante, si las reglas establecidas en las legislaciones no fueran lo suficientemente claras o completas (o incluso, inexistentes), pueden tomarse como fundamentos los estándares internacionales sobre privacidad y seguridad de los datos, en hardware y software.

En este artículo, se presenta una línea de investigación que aborda el *Análisis y Resolución de los Problemas Asociados al Diseño de Sistemas de IoT*.

Palabras clave: *IoT, Dispositivo, Cosa, Escalabilidad, Seguridad, Privacidad, Arquitectura, Internet.*

CONTEXTO

La presente línea de investigación se enmarca en el proyecto de investigación denominado *"Ingeniería del Software: Estrategias de Desarrollo, Mantenimiento y Migración de Sistemas en la Nube"*. El proyecto, a su vez, es un *Proyecto de Investigación Consolidado - PROICO*, y posee el código *03-2020*. Tal proyecto es la continuación de diferentes proyectos de investigación, a través de los cuales se ha logrado un importante vínculo con distintas

universidades a nivel nacional e internacional.

1. INTRODUCCIÓN

El Internet Of Things (IoT) comprende sistemas que van desde dispositivos individuales transmitiendo datos hacia un servidor en la nube hasta complejos sistemas industriales conformados por arquitecturas de múltiples capas de dispositivos y servidores, involucrando comunicaciones en una, dos o más direcciones [5, 6]. A su vez, cada componente de la arquitectura del sistema puede ubicarse en diferentes regiones geográficas y utilizar diferentes tecnologías, por lo que su intercomunicación podría requerir de múltiples medios de conexión física, y de la traducción entre diferentes protocolos de comunicación.

La estructura de costos asociada a la comunicación podría variar notablemente de una región geográfica a la otra por causas como las presentadas a continuación:

- Costos de acceder a fuentes de energía eléctrica.
- Medios de conexión a Internet disponibles. Dependiendo del medio físico utilizado, la cantidad de usuarios conectados en simultáneo, las distancias recorridas por la información y los impuestos estatales, podrían variar notablemente los costos. A su vez, podría ser que no existan medios de conexión a Internet disponibles en la ubicación de despliegue de los dispositivos, por lo que los mismos deberían ser conectados mediante otro medio de comunicación a una Puerta de Enlace (Gateway), que luego los conecte a Internet.
- Los proveedores de servicios utilizados en el sistema de IoT podrían cobrar tasas diferenciales de acuerdo a la ubicación geográfica desde la cual se acceda a los mismos [11].

A su vez, existen una serie de características deseadas en un sistema de IoT que podrían influir notablemente en su diseño, y que serán mencionadas a continuación [3, 5, 7, 8, 9]:

- Una de las características deseadas en un sistema de IoT es la privacidad en las comunicaciones, para asegurar que los mensajes transmitidos, únicamente puedan ser comprendidos por los receptores indicados al momento de iniciar las comunicaciones.
- Otra característica importante, y que está asociada a la privacidad, es la robustez de las comunicaciones. Esta implica asegurar la inmutabilidad de los mensajes transmitidos, desde su emisión hasta su recepción, y garantiza que los datos no serán tergiversados por fallas en los dispositivos, en los medios de comunicación, o por personas con intenciones maliciosas. La robustez es esencial en sistemas críticos que requieren de información precisa para evitar fallas catastróficas, como podrían ser las plantas de energía nuclear y la tecnología de uso militar.
- La eficiencia de los dispositivos podría ser de gran interés para asegurar largos períodos de funcionamiento en regiones donde se disponga de una red eléctrica inestable. De igual forma, la eficiencia de los dispositivos podría favorecer notablemente a la reducción de costos asociados al consumo de energía, especialmente en aquellos sistemas donde se deban desplegar grandes cantidades de dispositivos, y donde los costos de acceso a fuentes de energía sean elevados.
- La precisión y velocidad de las mediciones realizadas por los sensores o de las acciones realizadas por los actuadores, tienen una gran influencia en los costos de los dispositivos.
- El nivel de resistencia de los componentes electrónicos que conforman los dispositivos frente a las inestabilidades del ambiente en el que se despliegan, varía notablemente los costos de los mismos.
- La escalabilidad de los modelos de comunicación usados es también muy

importante, a nivel horizontal o vertical. A nivel horizontal, un diseño arquitectónico escalable favorece la instalación de grandes cantidades de dispositivos. A nivel vertical, un diseño arquitectónico escalable permite la conexión de diferentes clases de dispositivos sin que se presenten mayores dificultades.

- Es de gran importancia tener en cuenta los tiempos de respuesta deseados (i.e. la latencia deseada), especialmente en sistemas críticos que dependan de información en tiempo real, como podrían ser aquellos instalados en plantas de producción de energía nuclear o de productos químicos de alto riesgo.
- Por último, la selección del protocolo de comunicación a utilizar en cada parte del sistema de IoT influye notablemente en su diseño y costos. Un protocolo de comunicación muy complejo puede requerir que se incrementen las capacidades del hardware de los dispositivos, al mismo tiempo que se incrementan los tiempos de investigación y desarrollo [12].

Como se puede imaginar el lector, no pueden maximizarse todas las características mencionadas, sin afectarse entre ellas:

- Si se desea incrementar la privacidad y robustez en las comunicaciones, es probable que deban implementarse protocolos de comunicación de mayor complejidad. Sin embargo, también será necesario incrementar las capacidades de hardware de los dispositivos para poder ejecutarlo sin inconvenientes, lo cual incrementa los costos y disminuye la eficiencia energética de los mismos.
- De forma similar, si se desea garantizar tiempos de respuesta muy pequeños utilizando componentes electrónicos de bajo costo, es probable que la robustez en las comunicaciones se reduzca, la precisión de los sensores disminuya y los datos deban transmitirse múltiples veces

para tener una mayor confiabilidad. Además, los dispositivos de menor costo podrían ser más sensibles a los cambios en el ambiente.

- Si se utilizan protocolos de red muy complejos y seguros, y los dispositivos no poseen la capacidad necesaria para manejarlos, los medios de conexión son inestables, o bien, los servidores Web se encuentran saturados, la latencia podría presentar grandes fluctuaciones en el tiempo y poner en riesgo el correcto funcionamiento de sistemas críticos, al mismo tiempo que la vida de las personas involucradas.

Hasta el momento, se han mostrado los diferentes desafíos técnicos y comerciales que se pueden presentar al momento de diseñar un sistema de IoT. Sin embargo, existen otras cuestiones a tener en cuenta, dependiendo de quien sea la entidad interesada en la creación del sistema:

- Si la entidad interesada pertenece a algún gobierno, es muy probable que se prioricen componentes de hardware o software de producción local. Esto mismo ocurriría con los servicios contratados. Además, es posible que la información recopilada deba ser publicada en algún repositorio de acceso público, por lo que seguramente se busque preservar el anonimato de las personas asociadas a la misma. Para evitar conflictos, los sistemas deberán diseñarse con el cuidado de cumplir con las leyes vigentes y con acuerdos internacionales a los que el gobierno se encuentre suscrito [1, 2].
- Si la entidad es privada, el nivel de privacidad de la información estará sujeto a contratos aceptados por los clientes. No obstante, en sistemas ubicados en ámbitos públicos podría recopilarse información de personas que desconozcan la existencia o fines del sistema de IoT, por lo cual deberán tratarse esos datos de acuerdo a las

legislaciones vigentes y teniendo en cuenta los estándares internacionales [4, 5].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

A partir de lo mencionado en la sección anterior, puede verse que el correcto diseño e implementación de un sistema de IoT puede ser un desafío difícil de superar si se lo intenta abordar en su totalidad. Lo mejor es identificar cada uno de los problemas involucrados y atacarlos con la mejor solución posible, de acuerdo a las tecnologías y conocimientos del estado del arte. Además es necesario detallar cuáles son los resultados que se espera que ofrezca el sistema resultante, con el fin de establecer límites sobre los cuales buscar soluciones a los problemas que se presenten.

Con el fin de facilitar la búsqueda de soluciones óptimas a los problemas que se presenten en el diseño de un sistema de IoT escalable y seguro, esta línea de investigación propone:

- Investigar en detalle los problemas asociados al despliegue de dispositivos de IoT en diferentes regiones del planeta, y sus posibles soluciones.
- Investigar en detalle la estructura de costos que conforma la transferencia de datos entre diferentes regiones del planeta.
- Investigar en detalle los diferentes factores que influyen en la latencia de las comunicaciones de los sistemas de IoT, y el estado del arte de las tecnologías y metodologías utilizadas para tener un mayor control sobre estos tiempos.
- Investigar el estado del arte de los patrones arquitectónicos usados en sistemas de IoT de diferentes escalas, que faciliten la escalabilidad y seguridad de los mismos.

3. RESULTADOS OBTENIDOS/ESPERADOS

A partir de las investigaciones realizadas se espera poder concretar los siguientes objetivos:

En principio, se realizarán informes que detallarán los costos asociados a la transferencia de datos entre componentes de un sistema de IoT, dependiendo de su ubicación geográfica, el volumen de datos transmitido, la periodicidad de la transmisión y el protocolo de comunicación utilizado.

De igual forma, se realizarán informes sobre la latencia en las comunicaciones de sistemas de IoT, los cuales permitirán realizar estimaciones, detectar cuellos de botella en las comunicaciones y encontrar formas para eliminarlos y reducir la latencia.

A partir de los informes realizados, se creará un sistema de IoT en el cual se buscará optimizar las comunicaciones y costos asociados.

4. FORMACION DE RECURSOS HUMANOS

Los progresos obtenidos en esta línea de investigación sirven como base para el desarrollo de tesis de posgrado, ya sea de doctorado o maestrías en Ingeniería de Software y desarrollo de trabajos finales de las carreras Licenciatura en Ciencias de la Computación, Ingeniería en Informática e Ingeniería en Computación de la Universidad Nacional de San Luis, en el marco de los Proyectos de Investigación mencionados en la Sección *Contexto*.

5. BIBLIOGRAFÍA

- [1] E. U. (EU); What is GDPR, the EU's new data protection law?, 2018.
- [2] E. S. y. C. d. D. d. I. N. Argentina; Ley 25.326, de la Protección de los Datos Personales, 2000
- [3] Microsoft; Azure IoT Signals; 14/10/2021; <https://azure.microsoft.com/es-es/resources/iot-signals/>
- [4] ETSI; TS 103 645; Cyber Security for Consumer Internet of Things, 2019
- [5] IEEE; The Institute, Internet Of Things; 2014;

- https://iot.ieee.org/images/files/pdf/The_Institute-IoT.pdf
- [6] SAP; What is Internet Of Things; 2022; <https://www.sap.com/latinamerica/insights/what-is-iot-internet-of-things.html>
- [7] T. Choudhury, A. Gupta, S. Pradhan, P. Kumar y Y. S. Rathore; Privacy and Security of Cloud-Based Internet of Things (IoT); 28/10/2017; <http://ieeexplore.ieee.org/document/8307328/>.
- [8] H. Ren, H. Li, Y. Dai, K. Yang y X. Lin; Querying in Internet of Things with Privacy Preserving: Challenges, Solutions and Opportunities; 13/03/2018; <http://ieeexplore.ieee.org/document/8315210/>.
- [9] IETF; The Internet of Things - Concept and Problem Statement; 2010; <https://www.ietf.org/archive/id/draft-lee-iot-problem-statement-05.txt>
- [10] A. Bassi, M. Bauer, M. Fiedler, T. Kramp, R. v. Kranenburg, S. Lange y S. Messner; Enabling Things to Talk: Designing IoT Solutions with the IoT Architectural Reference Model; *SpringerOpen*; 2013; <https://link.springer.com/content/pdf/10.1007%2F978-3-642-40403-0.pdf>
- [11] Microsoft; Azure Precios de Ancho de Banda; 2022; <https://azure.microsoft.com/es-es/pricing/details/bandwidth/>
- [12] Microsoft; IoT Technology Protocols; 2022; <https://azure.microsoft.com/en-us/overview/internet-of-things-iot/iot-technology-protocols/>

Marco de Referencia para la integración de Accesibilidad en Sistemas e-Learning

Iván Balmaceda Castro¹, Carlos Salgado², Mario Peralta², Alberto Sanchez²

¹Departamento Académico de Ciencias Exactas, Físicas y Naturales

Universidad Nacional de La Rioja, La Rioja, Argentina.

ibalmaceda@unlar.edu.ar

²Departamento de Informática. Facultad de Ciencias Físico-Matemáticas y Naturales

Universidad Nacional de San Luis, San Luis, Argentina.

{csalgado, mperalta, alfanego}@unsl.edu.ar

RESUMEN

En estos últimos años, la educación atraviesa una constante evolución, donde las nuevas tecnologías de información y comunicación se ven reflejadas en las actividades académicas a través de la inclusión de cursos en modalidad virtual en el proceso formativo como una gran alternativa en los procesos de enseñanza-aprendizaje. Sin embargo, no se toma en cuenta factores de accesibilidad que son fundamentales para que la mayor cantidad de personas puedan hacer uso de un sistema/producto/servicio.

La Línea de investigación, plantea la importancia de incorporar condiciones de accesibilidad en los entornos virtuales de aprendizaje, donde se propone un marco metodológico de referencia para tomar en cuenta aspectos de accesibilidad en los diferentes procesos del ciclo de vida de un proyecto educativo virtual.

Palabras clave: Marco de Referencia, Accesibilidad, e-Learning

CONTEXTO

Este trabajo se enmarca dentro del proyecto de investigación “Ingeniería de Software Estrategias de Desarrollo, Mantenimiento y Migración de Sistemas en la Nube”, Proyecto PROICO 03-2020 del Departamento de Informática de la Facultad de Ciencias Físico-Matemáticas y Naturales, de la Universidad Nacional de San Luis. Dicho proyecto se

encuentra reconocido por el programa de Incentivos. Y del proyecto de investigación “Modelo de Proceso y Evaluación Centrado en el Usuario incorporando requisitos de Usabilidad, Accesibilidad y Experiencia de Usuarios” Cod. Proy. 27PIN/C0003 del Centro de Investigación y Desarrollo Informático, Departamento Académico de Ciencias Exactas, Físicas y Naturales - Sede Regional Chamental Universidad Nacional de La Rioja.

1. INTRODUCCIÓN

La accesibilidad en los sistemas web en la presente sociedad de la información, tiene una repercusión a todo nivel, especialmente en la educación virtual.

Por ello, es importante identificar que en la actualidad la formación virtual o enseñanza basada en e-Learning, y que por el contexto sanitario generado por la Pandemia COVID-19, ha tenido un crecimiento cada vez mayor, no solo en las instituciones de educación superior, sino también en otras entidades educativas como la escuela en su nivel inicial, primario y secundario, las orientadas al aprendizaje en el trabajo o formación profesional. Actualmente, es posible identificar no solo plataformas educativas digitales, sino también, plataformas de cursos masivos en línea (MOOC) y en general nuevos espacios virtuales que se encuentran presentes en instituciones de educación a todo nivel.

En e-Learning, se utilizan entornos virtuales de aprendizaje que, mediante la utilización de

herramientas e instrumentos, los usuarios implicados, generan una interacción en el proceso de enseñanza-aprendizaje [1]. En este sentido, así como las limitaciones físicas influyen en la institución, las plataformas virtuales de aprendizaje, los contenidos curriculares pueden presentar también limitaciones a las personas con discapacidad.

La discapacidad, en los últimos tiempos, ha cobrado una gran significación que concierne a diferentes disciplinas encargadas de diseñar espacios donde una persona con discapacidad pueda acceder, por ejemplo, a la información con un alto grado de inclusión, que permita prever limitaciones que se les puedan presentar.

Más allá de esta significación, aún existe la falta de interés del profesional informático en crear sistemas universales y accesibles, dando un alto grado de prioridad a la parte estética y dejando muchas veces de lado los conceptos concernientes a la usabilidad y la accesibilidad, Lasso Guerrero, sostiene que, actualmente en la web se hace uso excesivo de recursos visuales, mientras que la usabilidad se ha visto reemplazada por herramientas que muchas veces pueden ser innecesarias [2]

Los problemas de accesibilidad y utilidad de los sistemas se han convertido en críticos, la tecnología se ha diversificado y cada vez es más necesario utilizar la tecnología para participar plenamente en situaciones cotidianas. Tanto la usabilidad como la accesibilidad debe ser entendida en toda su amplitud, fomentando en la creación de espacios inclusivos donde todas las personas puedan relacionarse de forma sencilla, sabiendo que las medidas a diseñar y desarrollar han de ser evaluadas desde la comprensión de la discapacidad como punto de encuentro entre el entorno, que desempeña un papel fundamental, y las características personales del individuo.[3]

Las necesidades de accesibilidad de los usuarios varían a lo largo del tiempo y en los distintos contextos de uso [1]. En esto, se torna

importante la calidad de los sistemas e-learning, para de esta forma asegurar un contenido de calidad y que se ajuste con los requerimientos para su correcto funcionamiento, pero por sobre todo que su interacción sea con la mayor cantidad usuarios, posean una discapacidad o no.

El ciclo de vida de un proyecto de e-learning se divide en diferentes procesos, conceptos, productos y stakeholders involucrados en las actividades educativas. Generalmente, los procesos que componen el ciclo de vida de un proyecto educativo virtual suelen estar basados en experiencias empíricas de las instituciones educativas.

Para ello, la Organización Internacional de Normalización propuso la norma ISO/IEC 40180 que proporciona los fundamentos y el marco de referencia para el aseguramiento de la calidad, la gestión de la calidad y la mejora de la calidad en el aprendizaje, la educación y la formación mejorados en TI (denominado E-Learning). [4]

Esta norma, consiste principalmente en el Quality Reference Framework (QRF) para E-Learning [5], este es un marco común y genérico que describe, especifica y comprende propiedades, características y métricas de calidad.

En [4] define que “el QRF combina un modelo de proceso elaborado y extenso con un modelo descriptivo de los procesos. ISO/IEC 40180 armoniza los enfoques, conceptos, especificaciones, términos y definiciones existentes relacionados con la calidad para el aprendizaje electrónico, la educación y la formación.”

El modelo de proceso de referencia de esta norma internacional ISO 40180 proporciona un marco general para diseñar la estructura para el aprendizaje, la educación y la formación basado en el consenso global, compuesto por:

- Análisis de Necesidades

- Análisis del Marco
- Concepción y Diseño
- Desarrollo y Producción
- Implementación
- Proceso de Aprendizaje
- Evaluación

Cada uno de los procesos contiene una serie de subprocesos para un total de 38 sub-procesos.

Por su parte, la Norma ISO 19796-3 [6], presenta una propuesta para describir las técnicas y procedimientos involucrados en los procesos de un ciclo de vida de un proyecto educativo virtual. En esta misma norma, también se establece cómo describir los productos que se obtienen de las técnicas utilizadas, y define las métricas para verificar la calidad de los productos relacionados con cada fase del proceso de aprendizaje, pero no entra en especificidades sobre la accesibilidad, sino más bien la propuesta es generalista.

En este sentido, la Norma ISO 40180 [4] es un punto de partida para incorporar los aspectos de accesibilidad a cada uno de los procesos y componentes de un proyecto educativo basado en modalidad e-learning, teniendo en cuenta las técnicas y procedimientos de [6].

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Esta línea de investigación se centra en un enfoque holístico para la realización de proyectos educativos virtuales accesibles, teniendo en cuenta diferentes aspectos de accesibilidad que intervienen en los procesos de creación de plataformas virtuales.

La accesibilidad del e-learning es una tarea compleja que implica un esfuerzo multidisciplinar principalmente desde el punto de vista tecnológico, didáctico y administrativo. En este sentido, una plataforma de e-learning debe ser accesible, pero la parte más importante es el contenido de e-learning para tener una solución efectiva.

3. RESULTADOS ESPERADOS/OBTENIDOS

La línea de investigación presentada tiene como objetivo general: Proponer un Marco de Referencia que integra un conjunto de criterios, posibles problemas y soluciones, a considerar, para mejorar la de las plataformas virtuales de aprendizaje. Para el mismo, se plantean los siguientes objetivos particulares:

- Definir el estado del arte en la temática relacionada a la accesibilidad en los procesos de un proyecto educativo virtual .
- Analizar los estándares de accesibilidad web relacionados con plataformas de e-learning.
- Proponer una solución metodológica que incorpore aspectos de accesibilidad en los procesos.
- Definir y validar la propuesta por medio de los Campus Virtual de la UNLaR y la UNSL.

Como marco de referencia para describir los procesos y componentes del ciclo de vida de un proyecto de e-learning, se tendrá en cuenta el Modelo de Proceso de referencia de ISO/IEC 40180, donde se añadirán las características de accesibilidad a los componentes siguiendo las mejores prácticas y las directrices presentadas en la norma.

El marco propuesto detallara las adaptaciones para incorporar la accesibilidad en los diferentes procesos de producción de Plataformas virtuales de aprendizaje. Este, trata un modelo de proceso de siete partes dentro del ciclo de vida. Al tratarse de un marco de accesibilidad basado en un estándar internacional, puede ser utilizado por cualquier institución educativa para describir, comparar y adaptar sus propios procesos de producción hacia una educación virtual accesible.

4. FORMACIÓN DE RECURSOS HUMANOS

La línea de investigación se da en un trabajo en conjunto entre la Universidad Nacional de La Rioja (UNLaR) y la Universidad Nacional de San Luis (UNSL).

En dicha línea, se está trabajando en el Proyecto de Tesis de la Maestría en Ingeniería de Software (Plan Ord. 05/2010-CD) del Lic. Iván Balmaceda Castro, Docente del Departamento Académico de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de La Rioja. Como así también se están llevando a cabo algunas tesinas de grado para la Licenciatura en Computación. Y trabajos finales de carrera de la Ingeniería en Informática e Ingeniería en Computación de la Facultad de Ciencias Físico-Matemáticas y Naturales. Del mismo modo, Trabajos Finales de la Tecnicatura en Informática, Analista Universitario en Sistemas y Licenciatura en Sistemas de Información de la Universidad Nacional de La Rioja

5. BIBLIOGRAFÍA

1. Balmaceda Castro, I., Salgado, CH, Peralta, M., Sánchez, A., Fernández, MA, & Vera, CE (2021). Hacia un modelo integral de Accesibilidad en Sistemas e-Learning. En XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021, Chilecito, La Rioja) .
2. Lasso Guerrero, J. (2013) Ergonomía en el diseño web: Usabilidad de sitios web dedicados al comercio electrónico en Buenos Aires. pp.5,47,9 Recuperado de https://www.palermo.edu/dyc/maestria_diseno/pdf/tesis_completas/101-Lazzo-juan.pdf
3. Sánchez, A. (2013). Calidad y accesibilidad del elearning: encuentro y reflexión desde una perspectiva inclusiva. Revista Digital de Investigación Educativa. Recuperado de http://www.revistaconecta2.com.mx/archivos/revistas/revista8/8_4.pdf.
4. ISO 40180:2017 Information Technology – Quality for learning, education and training.
5. Stracke, C. M., Tan, E., Teixeira, A., Pinto, M., Vassiliadis, B., Kameas, A., Sgouropoulou, C., & Vidal, G. (2018). Quality Reference Framework (QRF) for the Quality of Massive Open Online Courses (MOOCs). Online available at www.mooc-quality.eu/QRF
6. ISO (2009) ISO/IEC 19796-3:2009, ITLET Quality management, assurance and metrics, Part 3: Reference methods and metrics.

ISS – Innovación en Sistemas de Software

Enfoques y Tendencias en el Desarrollo de Aplicaciones Móviles con Resiliencia.

Pablo Thomas , Lisandro Delia , Leonardo Corbalán ,

Juan Fernández Sosa , Fernando Tesone , Verónica Aguirre , Verena Olsowy ,

Patricia Pesado 

Instituto de Investigación en Informática LIDI (III-LIDI)

Facultad de Informática – Universidad Nacional de La Plata

50 y 120 La Plata Buenos Aires

Centro Asociado a la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

526 e/ 10 y 11 La Plata Buenos Aires

{pthomas, ldelia, corbalan, jfernandez, ftesone, vaguirre, volsowy, ppesado}@lidi.info.unlp.edu.ar

RESUMEN

Se presenta una línea de investigación y desarrollo centrada en el estudio de aspectos de Ingeniería de Software aplicados a las diferentes tendencias en el desarrollo de Aplicaciones Móviles con resiliencia.

Palabras claves: Plataformas para Dispositivos Móviles - Aplicaciones Nativas- Aplicaciones Híbridas – Aplicaciones Interpretadas – Aplicaciones por Compilación Cruzada – Aplicaciones Web Progresivas – Offline First. – Instant App - Resiliencia

CONTEXTO

Esta línea de Investigación forma parte del Proyecto “*Metodologías, técnicas y herramientas de Ingeniería de Software en escenarios híbridos. Mejora de proceso*”, en particular del subproyecto *Ingeniería de Software para escenarios híbridos* del Instituto de Investigación en Informática LIDI de la Facultad de Informática, acreditado por el Ministerio de Educación de la Nación.

Hay cooperación con Universidades de Argentina y se está trabajando con Universidades de Europa y Latinoamérica.

1. INTRODUCCIÓN

Los avances tecnológicos en la comunicación móvil, junto con los avances

en el campo de la computación han dado lugar a la computación móvil, que puede definirse como un entorno de cómputo con movilidad física que brinda al usuario la capacidad de acceder a datos, información u otros objetos lógicos desde cualquier dispositivo en cualquier red, al mismo tiempo que va moviéndose, cambiando su locación geográfica [1].

Las características específicas de la computación móvil plantean nuevos desafíos en el desarrollo de software para este tipo de dispositivos. La necesidad de tratar con diversos estándares, protocolos y tecnologías de red; las capacidades limitadas, no sólo en cuanto a procesamiento, sino también, en cuanto a consumo, tamaño físico y capacidad de almacenamiento de los dispositivos; las restricciones de tiempo impuestas por un mercado altamente dinámico y la existencia de distintas plataformas de hardware y software son sólo algunas de las dificultades a las que se enfrentan los desarrolladores en esta área.

Las aplicaciones se generan en un entorno dinámico e incierto. En su mayoría se trata de aplicaciones pequeñas, no críticas, destinadas a un gran número de usuarios finales que son liberadas en versiones rápidas para poder satisfacer las demandas del mercado. En otros casos las aplicaciones son de mayor tamaño, a veces

para brindar movilidad a una parte de la funcionalidad de un sistema más grande, mientras que otras veces son el único punto de interacción del sistema.

Algunas iniciativas como *offline first* proponen nuevas maneras de abordar el desarrollo de aplicaciones móviles tomando como requerimiento no funcional la posibilidad de seguir brindando un servicio incluso cuando el dispositivo pierde conectividad con la red móvil.

Todas las particularidades previamente mencionadas hacen que el desarrollo de software para dispositivos móviles difiera considerablemente del tradicional. Ello conduce a nuevas prácticas y metodologías que promueven el crecimiento de la Ingeniería de Software como disciplina, acompañando este proceso de desarrollo tecnológico.

La existencia de una enorme competencia en el mercado de las aplicaciones móviles, promovida en parte por la proliferación de dispositivos móviles y el número de negocios que están migrando sus servicios a Internet, revela la importancia que implica para las empresas el posicionamiento en este mercado. Para conseguirlo, es necesario reducir al máximo el tiempo de desarrollo de las aplicaciones y al mismo tiempo hacer que éstas se ejecuten en el mayor número de dispositivos posible.

Este propósito se ve obstaculizado por la excesiva fragmentación de hardware y software existente, originada por el alto número de dispositivos distintos, con sus propios sistemas operativos y plataformas de desarrollo.

La alternativa más costosa para hacer frente a esta problemática consiste en el desarrollo específico de aplicaciones de manera nativa en cada una de las plataformas existentes, utilizando entornos de desarrollo, lenguajes de programación y tecnologías propias de cada plataforma.

El desarrollo nativo de aplicaciones móviles posee ciertas ventajas como la posibilidad de acceder sin limitaciones a

todas las características del dispositivo (cámara, GPS, acelerómetro y agenda, entre otras), el alto rendimiento, interfaces gráficas consistentes con el resto de la plataforma, la posibilidad de trabajar sin acceso a Internet y de correr procesos en segundo plano. Estas aplicaciones pueden distribuirse a través de las tiendas en línea correspondientes. Sin embargo, el precio de todas estas ventajas es alto: no es posible reutilizar el código fuente entre plataformas diferentes, el esfuerzo se multiplica y se elevan los costos de desarrollo, actualización y distribución de nuevas versiones.

Como alternativa al desarrollo nativo en múltiples plataformas, se presentan diversos enfoques de desarrollo que procuran optimizar la relación costo/beneficio compartiendo la misma base de código fuente entre las versiones para las distintas plataformas [2]. La construcción de aplicaciones Web Móviles constituye un ejemplo representativo de este enfoque. Estas aplicaciones se diseñan para correr dentro de un navegador, se desarrollan con tecnologías web bien conocidas (HTML, CSS y JavaScript), no necesitan adecuarse a ningún entorno operativo móvil específico; su puesta en marcha es rápida y sencilla.

Las desventajas de las aplicaciones Web Móviles recaen sobre su rendimiento. Los tiempos de respuesta se dilatan afectados por la interacción cliente-servidor y las restricciones de seguridad impuestas a la ejecución de código por medio del navegador limitan el acceso a todas las capacidades del dispositivo. Además, la experiencia de usuario dista de ser similar a la de las aplicaciones nativas, resultando menos atractiva para el usuario final.

El enfoque híbrido es una alternativa que posee también la ventaja de estar basado en tecnologías web estándar (HTML, Javascript y CSS) pero, a diferencia del anterior no funciona dentro de un navegador, sino en un contenedor web especial con mayor acceso a las

características del dispositivo a través de una API específica.

Las aplicaciones híbridas permiten la reutilización de código en las distintas plataformas, el acceso al hardware del dispositivo, y la distribución a través de las tiendas de aplicaciones. Sin embargo, conservan algunas de las desventajas de las aplicaciones Web Móviles: la utilización de componentes no nativos en la interfaz perjudica la experiencia de usuario, y la ejecución se ve ralentizada por la carga asociada al contenedor web.

El enfoque interpretado se presenta como una alternativa en donde las aplicaciones son traducidas en su mayor parte a código nativo, mientras que el resto se interpreta en ejecución. Se implementan de forma independiente de las plataformas utilizando diversas tecnologías y lenguajes, tales como Javascript, Typescript y XML, entre otros.

La obtención de interfaces nativas constituye una de las principales ventajas de este tipo de aplicaciones, mientras que la definición de nuevas componentes suele tener un alto grado de complejidad debido a la necesidad de definir abstracciones compatibles con diferentes plataformas.

Finalmente, las aplicaciones generadas por compilación cruzada también constituyen un tipo de desarrollo multiplataforma. Estas aplicaciones se compilan de manera nativa creando una versión específica de alto rendimiento para cada plataforma destino.

Un nuevo concepto ha surgido en los últimos años denominado Aplicaciones Web Progresivas (PWA por sus siglas en inglés). Una PWA es una aplicación web que utiliza las últimas tecnologías disponibles en los navegadores para ofrecer en dispositivos móviles una experiencia lo más parecida posible a la de una aplicación nativa.

Los objetivos que persiguen las PWA son: lograr el mayor rendimiento posible en dispositivos móviles, que la aplicación cargue de manera casi instantánea, que la

interfaz de usuario se parezca lo máximo posible a una nativa, que se pueda trabajar sin conexión (*offline first*) y que se puedan enviar notificaciones a los usuarios, como en una aplicación nativa.

Una nueva tecnología, desarrollada recientemente por Google, denominada TWA (Trusted Web Activities) o Actividades Web de Confianza, permite integrar una PWA con una aplicación Android. Las TWA son ejecutadas desde un APK y distribuidas desde Google Play Store. Muestran en pantalla completa un navegador web dentro de una aplicación Android sin mostrar la interfaz del navegador.

Las TWA representan un punto de inflexión para los desarrolladores que ahora pueden distribuir sus PWA en Google Play Store que ha dejado de ser una tienda exclusiva de aplicaciones nativas.

Desde 2017 los desarrolladores de Android tienen una nueva opción para hacer llegar sus apps a los usuarios finales. El concepto de *Instant App* permite ejecutar una funcionalidad específica de una aplicación sin necesidad de instalar la App completa.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

- Enfoques de desarrollo de Aplicaciones para Dispositivos Móviles.
- Metodologías y Técnicas de la Ingeniería de Software y su aplicación en el desarrollo de software para dispositivos móviles.
- Aplicaciones Nativas en Android [3].
- Aplicaciones Nativas en iOS [3].
- Aplicaciones Web Móviles.
- Aplicaciones Móviles Híbridas (PhoneGap [4], Ionic [5]).
- Aplicaciones Móviles Interpretadas (Appcelerator Titanium [6], NativeScript [7]).

- Aplicaciones Móviles generadas por compilación cruzada (Xamarin [8], Corona [9]).
- PWA, Instant App y Offline First.
- TWA (Trusted Web Activities),
- Análisis y estudio comparativo de requerimientos no funcionales, tales como rendimiento, consumo de energía, tamaño de software, entre otros, en los distintos enfoques de Aplicaciones Móviles.
- Experiencia de usuario en Aplicaciones Móviles generadas con distintos enfoques de desarrollo.
- Internet de las Cosas (IoT), Internet de Todo (IoE) y aplicaciones de sensado móvil o sensado urbano.
- Aplicaciones Móviles y su utilización en el proyecto de *Smart City*.
- El rol de los dispositivos móviles en aplicaciones de *crowdsourcing* o colaboración abierta distribuida.
- El rol de las aplicaciones móviles como herramientas para hacer frente a la pandemia de COVID-19
- Aplicaciones Móviles resilientes. Escenarios adversos: conectividad inestable, pérdida de conexión, ralentización del tráfico de red, ciberataques. Estrategias: *offline first*, almacenamiento local, memoria cache, *retry pattern* (reintento) *throttling pattern* (estrangulamiento), lineamientos para el desarrollo de apps móviles resistentes a ciber-ataques

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados obtenidos/esperados se pueden resumir en:

- Se ha estudiado el impacto que tiene el enfoque de desarrollo elegido sobre el rendimiento de las aplicaciones construidas. Se consideraron las plataformas iOS y Android junto a varios *frameworks* de desarrollo

multiplataforma. Los resultados y conclusiones fueron publicados en [10] y [11].

- Se desarrolló "*Informática UNLP*", una aplicación móvil para la comunidad de la Facultad de Informática de la Universidad Nacional de La Plata [12]. Esta aplicación contribuye a la mejor comunicación alumno/docente. Entre otras funciones incluye una cartelera virtual ajustable a las preferencias de los alumnos y utiliza tecnología de realidad aumentada para proveer información sobre la ocupación de las aulas en tiempo real. Las primeras versiones fueron liberadas para Android pero actualmente también está disponible para iPhone y iPad.
- En [13] se estudiaron las implicaciones de la elección de un enfoque de desarrollo determinado sobre la eficiencia energética de las aplicaciones generadas. Se contemplaron aplicaciones con alta carga de procesamiento, reproducción de video y acceso a imágenes generadas con diversos enfoques de desarrollo.
- En [14] se han publicado los resultados de un estudio sobre el impacto que tienen 9 tecnologías distintas de desarrollo de aplicaciones móviles sobre 23 características de interés para la Ingeniería de Software.
- En [15] se estudió el modo en que distintos *frameworks* de desarrollo multiplataforma afectan el tamaño de la aplicación construida. Los resultados son relevantes dado que el espacio disponible se ha convertido en un recurso crítico para muchos usuarios.
- Se ha cuantificado el impacto que tiene la elección del *framework* de desarrollo sobre tres de los requerimientos no funcionales más demandados por los usuarios de dispositivos móviles: rendimiento, consumo de energía y uso de espacio de almacenamiento. Los resultados y conclusiones se publicaron

en [16] y constituyen una guía para los desarrolladores al momento de elegir el *framework* más adecuado a sus expectativas.

- Las Aplicaciones Web Progresivas o PWA, por sus siglas en inglés, fueron analizadas en [17]. Dada su estrecha relación con las aplicaciones web móviles fueron comparadas con estas últimas explicitando diferencias y similitudes. Se concluyeron una serie de ventajas y desventajas de un enfoque respecto del otro.
- Se desarrolló “InnovApp”, una PWA que permite realizar una visita virtual guiada a la muestra de Ciencia y Tecnología 2019 llevada a cabo en el Centro de Innovación y Transferencia Tecnológica (CIyTT) de la Facultad de Informática de la UNLP
- Se han estudiado distintas estrategias para enfrentar la pandemia de COVID-19 que incluyen tecnología móvil. Para ello se analizaron las aplicaciones promovidas por los gobiernos de 22 países distintos como herramientas tecnológicas para combatir la Pandemia. Los resultados y conclusiones se publicaron en [18].
- Se implementó un prototipo de sistema de sensado móvil participativo que permite al ciudadano monitorear la calidad del agua de red que llega a su domicilio y, en caso de anomalías, alertar a la comunidad y autoridades competentes [19].
- Se continúa con el desarrollo de nuevas características de "*Informática UNLP*", una aplicación multiplataforma en continua evolución y crecimiento.
- Se prosigue con el estudio y el análisis del alcance de la tecnología de sensado móvil en aplicaciones de Internet de las Cosas (IoT), Internet de Todo (IoE) y sensado urbano.
- Se continúa con el análisis sobre el rol de las Aplicaciones Móviles en el

contexto de las ciudades inteligentes, haciendo foco especialmente en aquellas que implementan la colaboración abierta distribuida (*crowdsourcing*).

- Se examinarán las distintas formas de persistencia de datos en aplicaciones móviles en función del enfoque/*framework* de desarrollo utilizado y como estrategia para mitigar la inestabilidad de las conexiones.
- Se estudiarán y analizarán distintas estrategias para alcanzar mayor nivel de resiliencia en las aplicaciones móviles. Se considerarán distintas circunstancias de adversidad.
- Se promoverá el avance sostenido y continuo de la formación de los miembros involucrados en esta línea de investigación.

4. FORMACIÓN DE RECURSOS HUMANOS

Los integrantes de esta línea de investigación dirigen Tesinas de Grado y Tesis de Postgrado en la Facultad de Informática, y Becarios III-LIDI en temas relacionados con este proyecto. Además, participan en el dictado de asignaturas/cursos de grado y postgrado de la Facultad de Informática de la UNLP.

5. BIBLIOGRAFÍA

1. Hongxing Li, Guochu Shou, Yihong Hu, Zhigang Guo. *Mobile Edge Computing: Progress and Challenges*. 2016 4th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud). Oxford UK.
2. Spyros Xanthopoulos, Stelios Xinogalos, *A Comparative Analysis of Cross-platform Development Approaches for Mobile Applications*, BCI' 2013, Greece
3. Tracy, K.W., *Mobile Application Development Experiences on Apple's iOS and Android OS*, Potentials, IEEE, 2012.
4. <http://phonegap.com/>

5. <https://ionicframework.com/>
6. <http://www.appcelerator.com/>
7. <https://www.nativescript.org/>
8. <http://xamarin.com/>
9. <https://coronalabs.com/>
10. Delia L., Galdamez N., Corbalan L., Thomas P., Pesado P., *Un Análisis comparativo de rendimiento en Aplicaciones Móviles Multiplataforma*, CACIC 2015, UNNOBA Junín, Octubre 2015.
11. Delia L., Galdamez N., Corbalan L., Thomas P., Pesado P. *Approaches to Mobile Application Development: Comparative Performance Analysis* SAI Computing Conference (SAI), 2017. Londres, del 18 al 20 de Julio de 2017.
12. Fernández Sosa J., Cuitiño A., Thomas P., Delia L., Caseres G., Corbalán L., Pesado P. *“Informática UNLP” la App de la Facultad de Informática*. CACIC 2017. La Plata, del 9 al 13 de Octubre de 2017.
13. L. Corbalan, J. Fernandez Sosa, A. Cuitiño, L. Delía, G. Cáseres, P. Thomas, P. Pesado. *“Development Frameworks for Mobile Devices: A Comparative Study about Energy Consumption”*. 5th IEEE/ACM International Conference on Mobile Software Engineering and Systems. Gothenburg, Sweden. 27 y 28 de Mayo del 2018.
14. L. Delia, P. Thomas, L. Corbalan, J. Fernandez Sosa, A. Cuitiño, G. Cáseres, P. Pesado. *“Development Approaches for Mobile Applications: Comparative Analysis of Features”* SAI - Computing Conference 2018. London, United Kingdom. 10 al 12 de Julio de 2018.
15. J. Fernández Sosa, P. Thomas, L. Delía, G. Cáseres, L. Corbalán, F. Tesone, A. Cuitiño, P. Pesado. *“Mobile Application Development Approaches: A Comparative Analysis on the Use of Storage Space”*. XV Workshop de Ingeniería de Software - XXIV Congreso Argentino de Ciencias de la Computación CACIC 2018. Tandil, Argentina. 8 al 12 de octubre de 2018.
16. Corbalán, L., Thomas, P., Delía, L., Cáseres, G., Sosa, J. F., Tesone, F., & Pesado, P. (2019, June). A Study of Non-Functional Requirements in Apps for Mobile Devices. In *Conference on Cloud Computing and Big Data* (pp. 125-136). Springer, Cham
17. Aguirre, V., Ortu, A., Delía, L. N., Thomas, P. J., Corbalán, L. C., Cáseres, G., & Pesado, P. M. (2019). PWA para unificar el desarrollo Desktop, Web y Mobile. In *XXV Congreso Argentino de Ciencias de la Computación (CACIC)(Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019)*.
18. Fernandez Sosa, J., Aguirre, V., Delía, L., Thomas, P., Corbalán, L., & Pesado, P. M. (2020). COVID-19: un análisis comparativo de Apps. In *XXVI Congreso Argentino de Ciencias de la Computación (CACIC) (Universidad Nacional de La Matanza, 5 al 9 de octubre de 2020)*.
19. Fernández Sosa, J. F., Aguirre, V., Corbalán, L. C., Delía, L. N., Thomas, P. J., & Pesado, P. M. (2021). Sensado móvil como estrategia de participación ciudadana en Ciudades Inteligentes. In *XXVII Congreso Argentino de Ciencias de la Computación (CACIC)(Modalidad virtual, 4 al 8 de octubre de 2021)*.

Diseño y desarrollo de Serious Games para la rehabilitación de pacientes neurológicos implementando VRPN para las comunicaciones entre las interfaces y los dispositivos.

Javier J. Rosenstein, Rodrigo Gonzalez, Juan Salvador Portugal, Julian Argañaraz, Nicolás Ignacio Zarate, Héctor Sebastián Salinas

Instituto de Investigaciones, Facultad de Informática y Diseño, Universidad Champagnat, Belgrano 721, 5501 Godoy Cruz, Mendoza, Mendoza, Argentina.

rosensteinjavier@uch.edu.ar, gonzalezrodrigo@uch.edu.ar, thejuasz@gmail.com, hiory01@gmail.com, nicolaszarate23@gmail.com, hdseba_uch@hotmail.com

RESUMEN

En el desarrollo de sistemas de realidad virtual uno de los inconvenientes que se encuentran es la comunicación entre las aplicaciones y los dispositivos de adquisición de datos. Ya sea por no disponer de un método de acceso en forma directa o por necesitar independencia entre ambos, es decir que las aplicaciones corran en una plataforma y los dispositivos en otras. Para lograr esta independencia y a su vez permitir la integración de todo el sistema de realidad virtual, es necesario la implementación de algún protocolo de comunicaciones que permita esta vinculación heterogénea en tiempo real. Los dispositivos generalmente están asociados a funciones o características de los individuos que los utilizan y se necesita integrar los movimientos que estos representan hacia una interfaz gráfica que permita la realimentación neuronal del paciente y de este modo lograr la mejora cognitiva (efecto de neurofeedback deseado). El presente trabajo trata del análisis e implementación del protocolo de comunicaciones VRPN (Virtual Reality Peripheral Network) entre las partes de un entorno multimedia donde interactúan la adquisición de movimientos del usuario y la representación visual en un escenario virtual que permita la retroalimentación al usuario en tiempo real logrando una experiencia interac-

tiva e inmersiva. Logrando de este modo la producción de un sistema de capacitación / rehabilitación o mejor llamado Serious Game. **Palabras clave:** VRPN, EOG, EEG, BCI, Serious Games, Mirror Neurons, Neurofeedback

CONTEXTO

El presente proyecto se desarrolla en el Instituto de Investigaciones de la Facultad de Informática y Diseño de la Universidad Champagnat (Godoy Cruz, Mendoza), en el marco de la Licenciatura en Sistemas de Información; en cooperación con el Laboratorio de I+D+i en Neurotecnologías de la empresa Neuromed Argentina S.A. (Godoy Cruz, Mendoza).

Este trabajo es parte del proyecto de investigación que dio inicio en Septiembre de 2020 denominado “Diseño y desarrollo de un prototipo de Serious Game destinado a la rehabilitación de problemas neurológicos implementando VRPN para la comunicación de la BCI”.

1. INTRODUCCIÓN

El objetivo principal de la presente línea de investigación consiste en el diseño y

desarrollo de una BCI (Brain Computer Interface) [1] [2], que permita interactuar entre las señales generadas por un paciente neurológico [3] y una interfaz de realidad virtual. Este sistema debe permitir lograr el principio de neurofeedback [4], o retroalimentación hacia el paciente. De este modo se podrán mejorar sus capacidades cognitivas correspondientes. Esto tiene aplicación directa en los tratamientos de rehabilitación en pacientes de patologías neurológicas y cognitivas, principalmente en niños, este enfoque es importante ya que está demostrado que es la mayor causa de discapacidad en niños, según estadísticas como las publicadas por el Indec de su último informe del 2018 (ver figura1), en argentina dentro de las principales discapacidades, la mental-cognitiva es la mayor de las discapacidades entre las edades de 6 a 14 años.

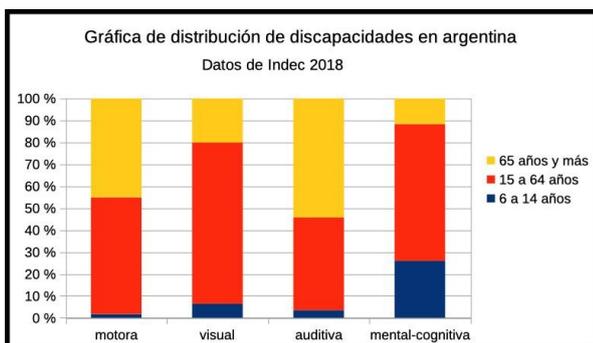


Figura 1: Discapacidades según Indec en argentina

Incluso, estudios indican la posibilidad de tratar patologías psiquiátricas como la depresión [5] [6].

La implementación del trabajo se organiza de acuerdo a las siguientes etapas:

- a) Adquisición de señales mediante técnicas de Electroencefalografía (EEG) [7] y adquisición de movimientos oculares mediante las técnicas de electrooculografía (EOG) [8] [9] [10] [11] [12].
- b) Análisis de estas señales en tiempo real para poder identificar la voluntad

de movimiento del individuo, así como la dirección del movimiento.

- c) Transformar la voluntad de movimiento en comandos del protocolo VRPN que permitan transmitir la información al componente software / hardware que la requiera.
- d) Como continuación al resultado obtenido en el punto anterior, Diseñar, programar y poder comandar una interfaz gráfica de aprendizaje o interfaz Cerebro / Computadora (BCI – Brain Computer Interface).

Finalmente se presenta la BCI como un sistema de adquisición de datos, procesamiento del protocolo serie a VRPN y luego la representación en nuestro modelo de prototipo de Serious Game [14] [15]. Una vez adquiridas las señales EEG/EOG, estas se analizan y codifican con las bibliotecas desarrolladas como parte de este proyecto. Esto permite su comunicación mediante VRPN con las interfaces virtuales que interpretan este protocolo.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

Este trabajo está compuesto por tres etapas o fases de trabajo. La primera de ellas corresponde a la implementación de VRPN entre un sistema simulado de captura de datos provenientes de un paciente, y una interfaz virtual básica. La segunda etapa del proyecto pretende avanzar sobre la captura de datos reales para que, luego de procesados, se transfieran a través de VRPN hacia la interfaz virtual de neurofeedback. Una breve descripción de cada una de las etapas se describe a continuación:

1. La primer parte consiste en la implementación de un simulador de señales EEG y EOG necesarias para el análisis e interpretación de la voluntad del usuario de la BCI. Estas señales una vez procesadas se deben codificar en comandos de VRPN para poder ser transmitidas hacia una interfaz virtual. Esta debe poder interpretar las señales transmitidas y representar la voluntad inicial del usuario correspondiente. De este modo se cumplen los objetivos de captura, análisis, procesamiento, transmisión, recepción y representación, lo cual produce el efecto de neurofeedback deseado sobre una interfaz virtual de capacitación a nivel prototipo.
2. La segunda etapa consiste en el diseño y desarrollo del escenario virtual de rehabilitación cognitiva para pacientes neurológicos según las indicaciones concretas por parte del especialista en neurología. Se partirá de un relevamiento de las técnicas de aprendizaje que se requieren implementar y los resultados que se pretenden obtener, indicados por el neurólogo o experto afín.
3. La etapa final futura de este proyecto consistiría en el desarrollo y mejora del proyecto mediante la adquisición real de señales EEG y EOG por electrodos ubicados superficialmente sobre la cabeza del paciente. Esta modificación al sistema requiere de un diseño e implementación electrónica así como del desarrollo del firmware que permita adquirir, analizar y

preprocesar estas señales, que luego serán transmitidas a través del protocolo VRPN hacia las interfaces virtuales de rehabilitación cognitiva de la etapa 2. Así, el sistema de adquisición real de datos se incorporará a futuro en forma transparente al proyecto implementado en las primeras 2 etapas.

3. RESULTADOS ESPERADOS

Los resultados esperados se pueden dividir en dos grandes grupos:

1. BCI e interfaz virtual gráfica que produzca en el paciente el efecto de neurofeedback: se espera obtener como producto final un sistema de retroalimentación a un usuario, en primera medida simulando la generación de señales que serán interpretadas en lops escenarios de realidad virtual diseñados para este fin, estas señales se deberán enviar mediante comandos codificados en el protocolo VRPN que se transmitirá a través de una red Ethernet.
2. Adquisición de datos reales de EEG y EOG: análisis y preprocesamiento; se identifican los comandos necesarios para su transmisión hacia el equipo generador de tramas VRPN para que finalmente se tenga un prototipo funcional que implemente la técnica de la terapia y se logre el efecto real final de neurofeedback propuesto, cabe aclarar que para que esto puedan lograrse es necesario la interacción con equipos de trabajo de diseño y desarrollo de electrónica que colaboren en conjunto con el equipo de investigación formado.

4. FORMACIÓN DE RECURSOS HUMANOS

La línea de I+D presentada está vinculada con el desarrollo de la tesis de maestría en teleinformática de la Universidad de Mendoza, de Javier J. Rosenstein, defendida a

finés del 2020. Dicha tesis se centró en la implementación del protocolo VRPN demostrando su uso en un sistema BCI.

Esta línea de I+D Además cuenta con la dirección del Mgter. Javier J. Rosenstein (UCH) y como investigador principal el Dr. Rodrigo Gonzalez (UCH). En lo que respecta a la formación de estudiantes de la licenciatura, esta línea de investigación cuenta con cuatro tesis de grado en curso, pertenecientes a los estudiantes Juan Salvador Portugal, Julian Argañaráz, Nicolás Ignacio Zarate y Héctor Sebastián Salinas, cuyos planes de tesis se encuentran específicamente dentro del marco de este proyecto. Todos ellos cursan actualmente la licenciatura en Sistemas de Información en la Universidad Champagnat.

Todos los avances logrados y las implementaciones realizadas relacionadas con el desarrollo del presente proyecto desde su primer etapa, se están utilizando como recurso para el dictado de talleres de comunicaciones, redes, programación de microcontroladores, programación en C/C++ y diseño y programación de interfaces virtuales de capacitación/rehabilitación en general, tanto para estudiantes de la universidad, como así también para alumnos externos a la UCH.

5. BIBLIOGRAFÍA

1. Y. Wang, X. Gao, B. Hong, and S. Gao, "Practical designs of brain-computer interfaces based on the modulation of EEG rhythms," in *Brain-Computer Interfaces*. Springer, 2009, pp. 137–154.
2. J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.
3. J. A. Pineda, "The functional significance of mu rhythms: translating "seeing" and "hearing" into "doing"," *Brain Research Reviews*, vol. 50, no. 1, pp. 57–68, 2005.
4. S. Enriquez-Geppert, R. J. Huster, and C. S. Herrmann, "Boosting brain functions: Improving executive functions with behavioral training, neurostimulation, and neurofeedback," *International Journal of Psychophysiology*, vol. 88, no. 1, pp. 1–16, 2013.
5. R. Ramirez, M. Palencia-Lefler, S. Giraldo, and Z. Vamvakousis, "Musical neurofeedback for treating depression in elderly people." *Frontiers in neuroscience*, vol. 9, pp. 354–354, 2014.
6. W. Rief, "Getting started with neurofeedback," 2006.
7. J. D. Kropotov, *Quantitative EEG, event-related potentials and neurotherapy*. Academic Press, 2010.
8. A. Bulling, J. A. Ward, H. Gellersen, and G. Troster, "Eye movement analysis for activity recognition using electrooculography," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 33, no. 4, pp. 741–753, 2011.
9. H. Singh and J. Singh, "A review on electrooculography," *International Journal of Advanced Engineering Technology*, vol. 3, no. 4, pp. 115–122, 2012.
10. D. P. Bautista, I. A. Badillo, D. De la Rosa Mejía, and A. H. H. Jiménez, "Interfaz humano-computadora basada en señales de electrooculografía para personas con discapacidad motriz," *ReCIBE*, vol. 3, no. 2, 2016.
11. S. Yathunathan, L. Chandrasena, A. Umakanthan, V. Vasuki, and S. Munasinghe, "Controlling a wheelchair by use of EOG signal," in *2008 4th International Conference on Information and Automation for Sustainability*, IEEE, 2008, pp. 283–288.
12. V. C. C. Roza, "Interface para tecnologia assistiva baseada em eletrooculografia," 2014.
13. A. C. Gaviria, I. C. Miller, S. O. Medina, and D. R. Gonzales, "Implementación de una interfaz hombre-computador basada en registros EOG mediante circuitos de señal mixta PSoC," in *V Latin American*

- Congress on Biomedical Engineering CLAIB 2011 May 16-21, 2011, Habana, Cuba. Springer, 2013, pp. 1194–1197.
14. P. Rego, P. M. Moreira, and L. P. Reis, “Serious games for rehabilitation: A survey and a classification towards a taxonomy,” in 5th Iberian Conference on Information Systems and Technologies. IEEE, 2010, pp. 1–6.
 15. J. S. Breuer and G. Bente, “Why so serious? On the relation of serious games and learning,” *Eludamos. Journal for Computer Game Culture*, vol. 4, no. 1, pp. 7–24, 2010.
 16. Instituto Nacional de Estadísticas y Censos, INDEC, Argentina, 2018, ISBN:978-950-896-519-6, https://www.indec.gob.ar/ftp/cuadros/poblacion/estudio_discapacidad_07_18.pdf
 17. Rosenstein, Javier J. (2020). *Uso de VRPN para la implementación de una interfaz cerebro-computadora* [Tesis de Maestría en Teleinformática, Universidad de Mendoza].

Línea de investigación realidad aumentada universal dirigida por Interacciones procedimentales en contextos 4.0

Martin Becerra ¹, Jorge Ierache ¹, María José Abasolo ^{2,3},

¹Universidad Nacional de La Matanza, DIIT, Grupo de Realidad Aumentada Aplicada

Florencio Varela 1903, La Matanza, Buenos Aires, Argentina

{mabecerra, jierache}@unlam.edu.ar

²Universidad Nacional de la Plata, Facultad de Informática, III-LIDI

³Comisión de Investigaciones Científicas de la Pcia. de Bs. As. CICIPBA

mjabella@lidi.info.unlp.edu.ar

Resumen

En este trabajo se enuncian la línea de investigación aplicada al desarrollo de un framework para navegadores de realidad aumentada semántico.

Palabras clave: Realidad Aumentada ubicua, Catálogo Virtual Aumentado, Ontologías, Web Semántica

Contexto

La investigación presentada es desarrollada por el grupo de investigación de Realidad Aumentada Aplicada del Departamento de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de La Matanza, como trabajo de tesis doctoral.

Introducción

La Realidad Aumentada (RA) permite la fusión de datos virtuales sobre el mundo físico, enriqueciendo con información virtual la percepción de la realidad [1]. En los últimos años, la RA se ha expandido a diferentes campos de aplicación tales como educación, salud, industria, turismo, marketing y entretenimiento. Nuestro equipo de investigación desarrolló diferentes aplicaciones, como juegos de tablero [2],[3], herramientas para la generación de materiales didácticos para el área educativa [4] y juegos didácticos [5], o sistemas de aumentación de información de salud mediante una tarjeta aumentada basadas en conocimiento para la

asistencia médica en emergencias [6],[7]. El proyecto de investigación aplicada se encuadra en el contexto de la aplicación de tecnologías de RA en la vida cotidiana de las personas que contribuye a la participación en el ámbito tecnológico generando así un impacto significativo en la sociedad.

En la actualidad existen diferentes navegadores de Realidad Aumentada (En inglés AR Browsers) populares en el mercado como LayAR[8], wiktitude[9] para proveer experiencias de Realidad Aumentada. Estas son limitadas ya que le permiten a un usuario consumir pasivamente un conjunto delimitado de funciones. Existen diferentes alternativas como ARCAMA3D [10], T. Matuszka et. al. [11] y SmartReality[12] que ofrecen una experiencia ubicua mediante la integración de tecnologías de web semántica para integrar información de la nube de datos interconectados (En inglés *Linked data Cloud*) para enriquecer las descripciones de puntos de interés cercanos a la posición de un usuario. Estas aplicaciones permiten crear contenidos, pero son consumidos estáticamente. En otras palabras, solamente pueden ver descripciones sin poder realizar ninguna acción sobre las mismas. Resulta de utilidad que el usuario pueda definir un procedimiento como conjunto de acciones a realizar en un entorno enriquecido por la Realidad Aumentada.

En la próxima sección se describe la línea de investigación y desarrollo “Realidad aumentada universal dirigida por Interacciones

procedimentales en contextos 4.0” que tiene como finalidad crear procedimientos para su explotación en contextos 4.0 en el que se apliquen tecnologías de Realidad Aumentada. El objetivo es permitir la creación de procedimientos accionables e interoperables mediante el uso de tecnologías de web semántica para dirigirnos hacia una Realidad Ubicua en donde el navegador de RA pueda consumir datos independientemente de la aplicación que los genere.

Líneas de Investigación, Desarrollo e Innovación

La línea de investigación y desarrollo realidad aumentada universal dirigida por Interacciones procedimentales en contextos 4.0 tiene por objetivo desarrollar aplicaciones de Realidad Aumentada en dirección al “Framework para la generación y explotación de procedimientos en Navegadores de realidad aumentada semánticos”.

El objetivo de esta línea de investigación aplicada se centra en el desarrollo de un framework[13] que permite a usuarios sin conocimientos específicos de programación, tener la capacidad de crear procedimientos o series de pasos a realizar en entornos físicos para su explotación con navegadores de Realidad Aumentada. En este prototipo se utiliza estándares de la web semántica para enriquecer dichos procedimientos con información de la web y lograr que sean interoperables gracias al uso de la ontología resultado de nuestra investigación que actúa como contrato semántico con otras aplicaciones convencionales como de RA.

Esta línea de investigación se puede aplicar en varios contextos que aplican tecnologías de Realidad Aumentada en especial en la industria 4.0 en la asistencia de tareas de trabajadores inteligentes. En hogares inteligentes se puede emplear para estructurar interacciones con dispositivos inteligentes de diferentes proveedores como también en la elaboración de

guías para el cuidado de personas en el área de salud.

El framework (figura 1) dispone de un editor llamado editor de procedimientos semánticos (en inglés Semantic Procedure Editor) que le permita armar procedimientos a un usuario creador de contenidos para articular tareas en un entorno real como por ejemplo la creación de tareas que debe realizar un operador inteligente en su puesto de trabajo en el contexto de la industria 4.0. Para la explotación de contenidos el prototipo dispone un navegador de Realidad Aumentada semántico (En inglés Semantic AR Browser) que permita al usuario buscar y utilizar dichos procedimientos. En primera instancia el prototipo se desarrolló para teléfonos móviles. El servicio que se encarga de unificar el acceso a los diferentes servicios del framework se lo llama en inglés Semantic Middleware, este es el responsable de comunicar el editor y el navegador con el servicio que administra el sistema de almacenamiento para publicar y explotar procedimientos semánticos para RA. (En inglés Public Semantic AR Procedure RDF triple store).

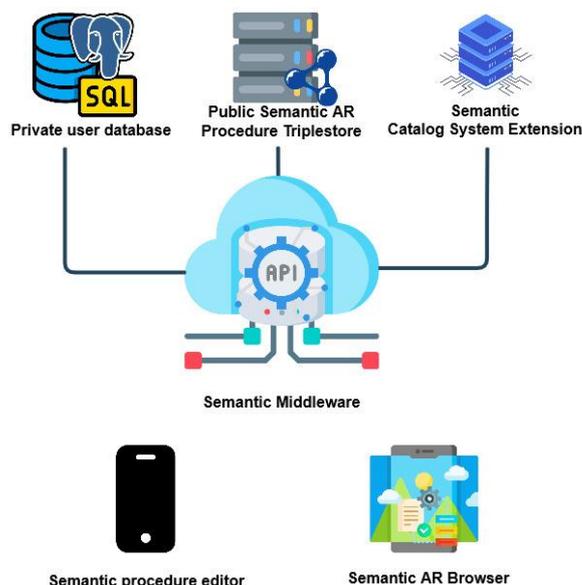


Figura 1 Esquema conceptual Framework para la generación y explotación de procedimientos en Navegadores de RA semánticos.

Se desarrolló el Sistema de Catálogos

Virtuales Aumentados [14], [15], [16] el cual permite la generación, distribución y explotación de contenidos de Realidad Aumentada. Dichos catálogos están compuestos por un conjunto de marcadores que son aumentados con información provista por los usuarios al momento de su creación texto, audio, video, modelos en 3D, la cual es visualizada utilizando una aplicación para teléfono inteligente conectada a internet.

En la línea presentada, la extensión del sistema de catálogos aumentada llamada en inglés Semantic Catalog System Extensión busca ser una capa semántica que permita consumir contenidos creados con este sistema, estructurando el acceso a datos mediante el uso de ontologías que permita que el sistema sea interoperable para que tome la función de repositorio de datos universal para todas aquellas aplicaciones de Realidad Aumentada que puedan interpretar el modelo ontológico proporcionado.

Resultados y Objetivos

En relación con la línea realidad aumentada universal dirigida por Interacciones procedimentales en contextos 4.0, se trabajó en la interacción procedimental para un navegador de Realidad Aumentada semántico [17] que tiene el objetivo de asistir a los usuarios en realizar tareas en su contexto. Un ejemplo que se puede citar es seguir una receta para cocinar un plato en el contexto de gastronomía 4.0. Esta interacción consiste en (1) Buscar un procedimiento (receta) a seguir, (2) Seleccionar uno entre los resultados obtenidos de las recetas a cocinar. (3) Seguir paso a paso de las acciones indicadas para alcanzar el objetivo como se puede observar en la figura 2.

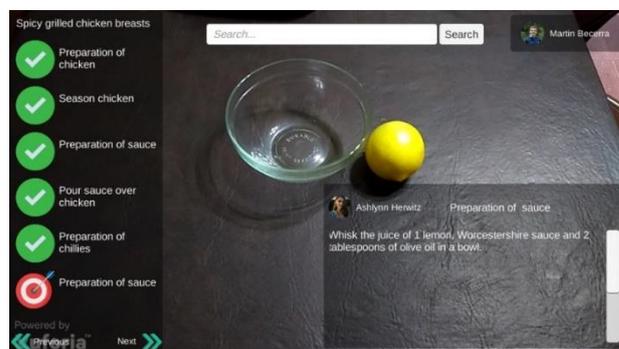


Figura 2 Ultimo paso de procedimiento para seguir una receta de cocina

Formación de Recursos Humanos

El grupo de investigación se encuentra conformado por 2 investigadores formados y dos investigadores en formación, trabajando en el área de RA. Uno de los investigadores en formación se encuentra realizando el Doctorado en Ciencias Informáticas en la UNLP, particularmente en el área específica del presente trabajo.

Referencias

- [1] Yee C., Abásolo M. J., Más Sansó R. y Vénere M. (2011). "Realidad Virtual y Realidad Aumentada. Interfaces avanzadas." ISBN 978-950-34-0765-3.
- [2] Ierache J., Mangiarua N., Verdicchio N., Sanz D., Montalvo C., Petrolo F., Igarza S. (2015). "Sistema de Catálogo Virtual Aumentado. Integración de Framework Especializado orientado a juegos didácticos" TEYET, pp 350-356, ISBN 978-950-656-154-3.
- [3] Verdicchio N., Sanz D., Igarza S., Mangiarua N., Montalvo C., Ierache J. (2016) "Sistema de Catálogo Virtual Aumentado Integración de Framework Especializado Orientado a Juegos Didácticos". TE&ET, pp 597-604, ISBN 978-987-3977-30-5.
- [4] Ierache J., Mangiarua N., Bevacqua S.,

Becerra M., Verdicchio N., Duarte N., Sanz D., Igarza S. (2014). "Herramienta de Realidad Aumentada para facilitar la enseñanza en contextos educativos mediante el uso de las TICs". *Revista Latinoamericana de Ing de Software*,1(1): -3, ISSN 2314-2642.

[5] Ierache J., Mangiarua N., Becerra M., Igarza S. Sposito O. Framework for the Development of Augmented Reality Applications Applied to Education Games. In: De Paolis L., Bourdot P. (eds) *Augmented Reality, Virtual Reality, and Computer Graphics. AVR 2018. Lecture Notes in Computer Science*, vol 10850. Springer, Cham. p. 340-350.

[6] Ierache N., Mangiarua N., Verdicchio D., Sanz D., Montalvo C., Petrolo F. and Igarza S., "Augmented. Card System Based on Knowledge for Medical Emergency Assistance". IEEE Xplore ISBN 978-1-5090-2938-9 2016.

[7] Ierache J., Verdicchio N., Duarte N., Montalvo C., Petrolo F., Sanz D., Mangiarua N., Igarza S., "Augmented Reality Card System for Emergency Medical Services", IWBBIO 2016 Proceedings Extended abstracts 20 - 22 abril 2016 Granada (SPAIN), pp 487-494, ISBN 978-84-16478-75-0.

[8] LayAR. Disponible en <https://www.layar.com/>. Accedido febrero 2021.

[9] Wikitude. Disponible en <https://www.wikitude.com/>. Accedido febrero 2021.

[10] Aydin B., Gensel J., Genoud P. Extending Augmented Reality Mobile Application with Structured Knowledge from the LOD Cloud. Disponible en https://www.researchgate.net/publication/278241470_Extending_Augmented_Reality_Mobile_Application_with_Structured_Knowledge_from_the_LOD_Cloud. Accedido febrero 2021.

[11] T. Matuszka et. al. The Design and Implementation of Semantic Web-Based

Architecture for Augmented Reality Browser. Disponible en https://www.researchgate.net/publication/280068074_The_Design_and_Implementation_of_Semantic_Web-Based_Architecture_for_Augmented_Reality_Browser. Accedido febrero 2021

[12] Nixon L., Grubert J. Reitmayr G. SmartReality: Integrating the Web into Augmented Reality. Disponible en https://www.researchgate.net/publication/257207430_SmartReality_Integrating_the_Web_into_Augmented_Reality. Accedido en febrero 2021.

[13] Becerra M., Ierache J., Abasolo M.J. (2021) Interoperable Dynamic Procedure Interactions on Semantic Augmented Reality Browsers. In: De Paolis L.T., Arpaia P., Bourdot P. (eds) *Augmented Reality, Virtual Reality, and Computer Graphics. AVR 2021. Lecture Notes in Computer Science*, vol 12980. Springer, Cham. https://doi.org/10.1007/978-3-030-87595-4_15.

[14] Ierache J., Mangiarua N., Verdicchio N., Becerra M., Duarte N., Igarza S. (2014). "Sistema de Catálogo para la Asistencia a la Creación, Publicación, Gestión y Explotación de Contenidos Multimedia y Aplicaciones de Realidad Aumentada". CACIC 2014 Red UNCI ISBN 978-987-3806-05.

[15] Ierache J., Mangiarua N., Bevacqua S., Verdicchio N., Becerra M., Sanz D., Sena M., Ortiz F., Duarte N., Igarza S. (2015). "Development of a Catalogs System for Augmented Reality Applications". World Academy of Science, Engineering and Technology, International Science Index 97, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 9(1), 1 - 7. ISSN 1307:6892.

[16] Mangiarua, N., Ierache, J.S., Becerra, M.E., Maurice, H., Igarza, S., & Sposito, O. (2018). Templates Framework for the Augmented

Catalog System. CACIC 2018 Red UNCI ISBN 978-3-030-20786-1.

[17] Becerra M., Ierache J., Abasolo M.J. (2022) Towards Augmented Reality Interactions driven by Universal Dynamic Procedural Browser actions in the 4.0 Contexts. In: Intelligent Systems Conference (IntelliSys) 2022 . Lecture Notes in Networks and Systems. Springer. En prensa.

Generación de un prototipo de comunicación sobre Android para app móviles reactivas

Susana Beatriz Chavez¹ Adriana Elizabeth Martín², Sergio Rafael Flores³; A. Sara Zogbe⁴, Ortiz, Alexis Rodrigo

Departamento e Instituto de Informática - F.C.E.F. y N. - U.N.S.J.
Complejo Islas Malvinas. Ignacio de la Roza y Meglioli.
C.P. 5402. Rivadavia. San Juan, 0264 4234129

¹schavez@iinfo.unsj.edu.ar; ²arianamartinsj@gmail.com;
³sergior@gmail.com; ⁴sarazogbe@yahoo.com.ar;
rodriunjs5@gmail.com

Resumen

El gran avance de las comunicaciones ha cambiado drásticamente la forma en que las personas y las máquinas interactúan entre sí, permitiendo el acceso instantáneo a información y servicios en tiempo real. El objetivo de este trabajo es proponer un modelo de comunicación entre app móviles, que permita evaluar la disponibilidad de los canales de comunicación y, de esta manera, garantizar que un mensaje llegue a destino. Para ello, se propone trabajar sobre una plataforma con soporte a la programación reactiva. Esto conlleva analizar y entender qué propone este nuevo paradigma reactivo para que el desarrollo del software móvil sea una solución real y competitiva.

Para poder interactuar continuamente con su entorno, las apps reactivas deben poder adaptarse a la carga a la que se enfrentan, utilizando una mayor capacidad computacional cuando sea necesario. Esto significa que debe poder hacer un uso eficiente del hardware en un solo dispositivo (que puede tener uno o más núcleos), y también ser capaz de funcionar a través de varios nodos de cómputo a su disposición, dependiendo de la carga.

Para garantizar la conectividad requerida por las apps reactivas, se pretende desarrollar un componente que identifique todas las alternativas disponibles de comunicación en el hardware del dispositivo. Considerando dispositivo a cualquier equipo o máquina que

sea capaz de generar y transmitir información a otro dispositivo.

Palabras claves: Paradigma Reactivo, Dispositivos móviles, IoT, Java, Android SO.

Contexto

El presente trabajo se encuadra dentro del área I/D Innovación en Sistemas de Software y se enmarca dentro del proyecto de investigación: Modelo de Sistema de Comunicación en Programación Reactiva, que ha sido aprobado por CICITCA y está en desarrollo para el período 2020-2022. Si bien este grupo de investigación se ha formado para trabajar en conjunto a partir del año 2020, cada integrante viene realizando tareas en distintos grupos, algunos desde enero de 2000 en tecnologías asociadas a la Computación Distribuida.

Se incorporaron dos alumnos en instancia de trabajo final de tesis de grado.

Las unidades ejecutoras para dicho proyecto son el Departamento e Instituto de Informática de la FCEfyN de la UNSJ.

Introducción

Debido a que los procesadores multinúcleo se están convirtiendo en un estándar, se han creado múltiples niveles de abstracción para simplificar la concurrencia y permitir un desarrollo más simple.

Es importante entender conceptos claves, como son los modelos de concurrencia [1]

utilizados para modelar y manipular operaciones asincrónicas sin perder de vista la tolerancia a fallos.

Cada uno de estos modelos tiene un enfoque diferente:

- **Futures:** son la base de la programación asíncrona y reactiva en Scala. Permiten manipular los resultados de una operación que aún no ha sucedido y lidiar efectivamente con la falla de dichas operaciones, permitiendo un uso más eficiente de los recursos computacionales.
- **Memoria transaccional de software (STM):** simplifica enormemente la comprensión conceptual de los programas multiproceso y ayuda a que los programas sean más fáciles de mantener al trabajar en armonía con las abstracciones de alto nivel existentes, como los objetos y los módulos. Usado por Clojure.
- **Stream Reactivos:** proporciona una abstracción para aplicaciones asíncronas altamente concurrentes con soporte para el procesamiento de flujo asíncrono con backpressure sin bloqueo. Ej.: RxJava, Reactor, etc.
- **Modelo basado en Actores:** es un modelo concurrente de cómputo para crear sistemas altamente concurrentes y paralelizables en un entorno distribuido. Popularizado por el lenguaje de programación Erlang y es implementado por Akka sobre JVM [2]. Con este modelo los actores no invocan a métodos, sino que se envían mensajes entre ellos (ver Figura 1). El envío de un mensaje no transfiere el hilo de ejecución al receptor. Un actor puede enviar un mensaje y continuar sin quedar bloqueado. Por lo tanto, puede lograr más en la misma cantidad de tiempo. A diferencia del comportamiento de los *objetos*, que libera el control de su hilo de ejecución cuando el método invocado termina (return). Sin embargo, los actores se

comportan como objetos, cuando reaccionan a los mensajes y devuelven la ejecución al terminar de procesar el mensaje actual. De esta manera, los actores logran realmente la ejecución que siempre se imaginó para los objetos.

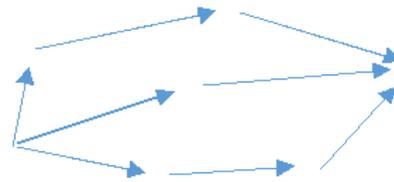


Figura1 . Interacción entre actores

La abstracción del modelo de actor permite pensar en el código en términos de comunicación.

No todos los modelos tienen un verdadero manejo de errores y recuperación de fallas. Algunos están diseñados para un alto rendimiento, pero no para escalar horizontal y/o verticalmente. Las ventajas que ofrecen cada uno de estos modelos pueden ser una “buena idea” en algunos contextos, pero no en otros. Es preciso cuestionar todo y asegurarse de que realmente se comprenden las fortalezas y debilidades de todas sus opciones de diseño, ya sea que se esté construyendo o no un sistema reactivo.

Protocolos de comunicación

Los dispositivos IoT (Internet of Things) utilizados tanto en entornos industriales, como en el entorno doméstico, pueden compartir aspectos de kernel (Linux, FreeRTOS, Windows Embedded, siendo el primero de ellos el más utilizado) y servicios de bajo nivel (*Real-Time Management, Context Discovery Management*), pero en términos de comunicación son muy diferentes. A continuación, se explican algunos de los protocolos utilizados a nivel doméstico e industrial.

Entorno doméstico

Cada día son más los hogares que disponen de uno o varios dispositivos IoT. como se muestra en la Figura 2. Existen multitud de

empresas que fabrican estos dispositivos. Cada uno tiene su propio protocolo de comunicación y aunque no es mucha la información que suelen facilitar, sobre todo si no son protocolos de código abierto, los más conocidos son: AllJoyn, HomePlug y HomeGrid, MFi (*Made For iPhone/iPod/iPad*), OCF (*Open Connectivity Foundation*), Thread (*network protocol*), etc.



Figura 2. Ejemplo IoT

Entorno industrial

Con la aparición de los dispositivos IoT, surge el concepto de industria 4.0 [10]. Podemos definirla como la digitalización completa a través de la integración de tecnologías de procesamiento de datos, software inteligente y sensores, desde los proveedores hasta los clientes. Pero para que esto sea posible, los dispositivos deben poder comunicarse, tanto entre sí, como hacia el exterior, como se muestra en la Figura 3. Algunos de los protocolos de comunicación existentes en la industria son: AMQP (*Advanced Message Queuing Protocol*), CoAP (*Constrained Application Protocol*), HTTP (REST/JSON) (*Hypertext Transfer Protocol*), XMPP (*Extensible Messaging and Presence Protocol*), etc.

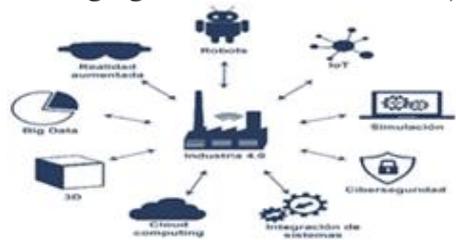


Figura3. Ejemplo IoT en la industria

Cualquier dispositivo móvil está preparado para acceder a internet de manera instantánea a información y servicios en tiempo real. Las especificaciones de los dispositivos Android, en la actualidad, son lo suficientemente competentes, con CPU de 4 a 10 núcleos de 3 a 4 GB de memoria. Se pueden usar hasta alrededor de 512 MB de memoria por máquina virtual (VM) con una configuración `largeHeap = true`.

Esta conectividad de los dispositivos con internet ha llevado a la creación de nuevos conceptos como Internet de las Cosas (IoT) que ha ganado popularidad en los últimos años [3].

Varias aplicaciones dependen de la potencia informática y otros recursos proporcionados por la nube para su correcto funcionamiento. Los dispositivos como sensores y dispositivos portátiles utilizan los servicios en la nube para procesar los datos que generan.

El IoT, tiene como objetivo contar con sensores u objetos dispersos para que generen información desde cualquier sitio accesible o bien en el interior de una máquina [4][5][6]. Esto llevará a un futuro en el que no sólo sea usado para la comunicación entre personas, sino, entre humano y máquina, e incluso, entre diferentes máquinas (M2M) [7]. Por ello, cobran también importancia los Smart Objects: objetos físicos con un sistema embebido que le permite procesar información y comunicarse con otros dispositivos y realizar acciones con base en una acción o evento determinado [8][9].

No obstante, todos estos sistemas complejos presentan un problema al momento de interconectarlos debido a las diferencias entre software y hardware utilizado en los diferentes procesos y canales de comunicación.

Los desafíos no solo involucran elementos como el almacenamiento de datos o el poder de la computación, sino que también exigen soluciones de software que puedan administrar y procesar la comunicación entre aplicaciones.

Este hecho pone de manifiesto la necesidad de nuevas formas de manejar la comunicación de datos que se originan en estos dispositivos.

Líneas de investigación, Desarrollo e Innovación

Para poder interactuar continuamente con su entorno, las apps reactivas deben poder adaptarse a la carga a la que se enfrentan, utilizando una mayor capacidad computacional cuando sea necesario. Esto significa que debe poder hacer un uso eficiente del hardware en un solo dispositivo (que puede tener uno o más núcleos), y también ser capaz de funcionar a través de varios nodos de cómputo a su disposición, dependiendo de la carga.

En pos de garantizar la conectividad requerida por las apps reactivas, se pretende desarrollar un componente de comunicación que identifique todas las alternativas disponibles de comunicación en el hardware del dispositivo. Se considera dispositivo a cualquier equipo o máquina que sea capaz de generar información y pueda transmitirla a otro dispositivo.

Las alternativas que cobran importancia se enfocan en las tecnologías inalámbricas como 4G, GSM y UMTS, Bluetooth y otras actualmente en desarrollo, particularmente las relativas a redes inalámbricas de área personal (WPANs). A estas, se suma el sistema de localización GPS, las medidas de tiempo de ultrasonido, UWB (Ultra- Wide Band), radiobalizas (por ej. vecinos lectores RFID con coordenadas conocidas o estaciones base WLAN) y tecnologías ópticas.

El objetivo que se persigue es lograr con éxito enviar un mensaje de un nodo a otro, siguiendo el *modelo de actores*. Pasar mensajes entre nodos, es la norma ahora, ya sea a través de dispositivos o computadoras en red.

En este modelo de concurrencia, ya no hay memoria real compartida, los diferentes

núcleos de un dispositivo se envían entre sí, fragmentos de datos explícitamente, tal como lo hacen las computadoras conectadas en una red. En lugar de ocultar el aspecto del paso del mensaje a través de variables marcadas como compartidas o utilizando estructuras de datos atómicas, un enfoque más disciplinado y basado en principios consiste en mantener el estado local a una entidad concurrente y propagar datos o eventos entre entidades concurrentes explícitamente a través de mensajes.

Los protocolos de IoT, tanto para el sector Doméstico como para el Industrial, están resueltos siempre y cuando haya conexión real de internet (datos o wifi). Cuando estas conexiones no están disponibles, no es posible mantener la comunicación, solo queda esperar a que se restablezca la misma, con las consecuencias que esto pudiera acarrear. Por este motivo se propone el diseño de un driver que se encargue de guiar en forma automática la comunicación de acuerdo a un esquema de jerarquía y en forma transparente al usuario.

En la Figura 4 se muestra cómo funciona el sistema de comunicación en forma normal.

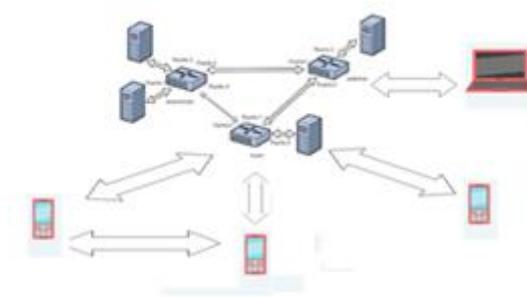


Figura 4. Comunicación normal

Cuando la comunicación se interrumpe debido a diversos factores, como por ejemplo fuera de cobertura, falta de crédito, o por cualquier otro motivo. Existen otras formas de comunicación que siguen vigentes sin ser diseñadas para este propósito, como por ejemplo el sistema de mensajería de texto o las llamadas de voz, incluso algunas

compañías proveen de servicio gratuito al sistema de comunicación WhatsApp, que si bien están destinados a un fin específico, se podrían utilizar para proveer de conectividad de manera temporal al móvil. De esa manera un móvil puede hacer de anfitrión para proveer la comunicación, por ejemplo: como se muestra en la Figura 5 el móvil M1 tiene comunicación con un sistema que a su vez se comunica con el móvil M3, si se rompe la comunicación de M1, el driver busca un anfitrión como por ejemplo el móvil M2 y accede a través de él al sistema

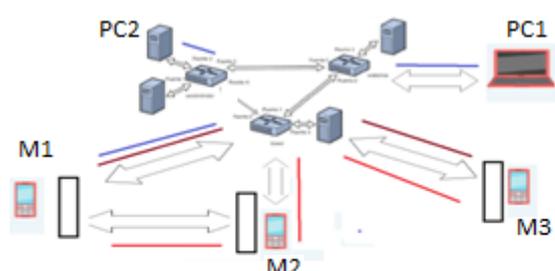


Figura 5. Comunicación a través de un anfitrión

Para ello es necesario la implementación de un driver que se comunique directamente con el sistema operativo y pueda discernir el mejor camino cuando se pierda la conexión natural de datos o wifi. La figura 6 muestra un esquema general de las funciones del driver.

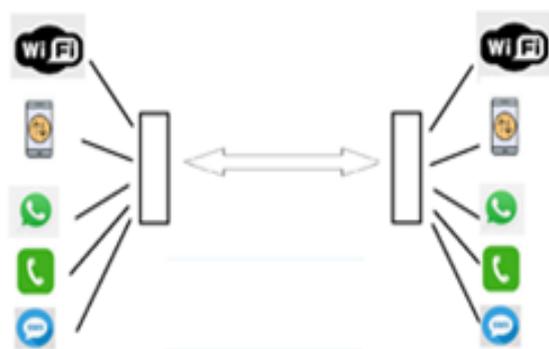


Figura 6. Funciones del Driver propuesto

Resultados y Objetivos

Resultados Obtenidos

Si bien, este equipo de investigadores se ha formado para trabajar en conjunto a partir del año 2020, cada integrante viene realizando tareas en distintos grupos, algunos desde enero de 2000 en tecnologías asociadas a la Computación Distribuida.

Con esta propuesta de trabajo se espera contribuir a la profundización y consolidación del conocimiento en esta área temática por parte de cada uno de los integrantes de este proyecto

Durante el periodo 2020-2021 transcurrido se realizaron diferentes publicaciones a fines a la propuesta.

A pesar de ser un periodo atípico debido a la pandemia el prototipo propuesto se encuentra en un estado avanzado de desarrollo e implementación.

Actualmente se está trabajando en establecer la comunicación punto a punto de los distintos sistemas de comunicación sin hacer eco con el dispositivo, es decir, generar mensajes internos, enviarlos, recibirlos en forma transparente al usuario, de modo que no los visualice y tampoco intervenga. El entorno de desarrollo es Android Studio, con el lenguaje de programación nativo java.

Paso seguido es comunicar las apps a través del canal de voz, es decir por medio de una llamada de voz. En este caso se deben realizar los siguientes pasos:

- Establecer la llamada con el móvil destino, se realiza la llamada el móvil destino responde
- Sincronizar la comunicación con el móvil destino, es decir, que sepa que no es una llamada de voz común, sino que se trata de una comunicación entre máquinas, se puede realizar esto por medio de una escucha de las notificaciones de llamadas, si es el número programado capturar la

- llamada, enviar un ok para sincronizar (en otra frecuencia).
- c) El texto a enviar se debe modular con una frecuencia de 42KHz, luego tomar los datos pasarlo a voz y modularlos.
 - d) Una vez sincronizado enviar la señal modulada.
 - e) En el dispositivo destino tomar la señal modulada con un formato de voz y aplicar el proceso de demodulación
 - f) El dato estaría disponible

La variedad de dispositivos con capacidades diferentes, es enorme y está al alcance de la mano. Resta configurar, adaptar, implementar y poner a punto un prototipo del modelo SiCo, por parte de este equipo de trabajo. De esta manera se contribuirá a la profundización y consolidación del conocimiento de esta área temática.

En cuanto a la movilidad de los dispositivos, los recursos pueden varían sin depender de los mismo, sino de los mecanismos de comunicación desde 4G pasando por GSM, solo WhatsApp, solo mensajería de texto o simplemente llamada de voz, es por ello que se necesita disponer de un mecanismo de selección automática del mejor sistema de comunicación disponible.

Objetivos

El objetivo del grupo de investigación es evaluar todas alternativas disponibles para manipular la comunicación entre aplicaciones que se originan en los distintos dispositivos, y construir un prototipo basado en una arquitectura reactiva donde la resiliencia y la comunicación asíncrona permitan que los sistemas estén exentos de errores.

Formación de Recursos Humanos

El equipo de trabajo está compuesto por 4 docentes-investigadores de la línea de investigación presentada que figuran en este trabajo y Alumnos avanzados de las Carreras

de Licenciatura en Sistemas de Información, Y Licenciatura en Ciencias de la Computación en estado de tesis; pertenecientes a la Universidad Nacional de San Juan.

Se está trabajando con dos tesinas en el ámbito de la programación para dispositivos móviles. Se espera iniciar otra tesina de grado en el área motivo de la presente propuesta de investigación.

Referencias

1. Adam L. Davis: Reactive Streams in Java_ Concurrency with RxJava, Reactor, and Akka Streams-Apress (2018)
2. How the Actor Model Meets the Needs of Modern, Distributed Systems: <https://doc.akka.io/docs/akka/current/typed/guide/actors-intro.html>
3. José I. Rodríguez M. Tesis Doctoral “Metamodelo para la integración del Internet de las cosas y Redes sociales” Universidad Oviedo 2017
4. González García Cristian, “MIDGAR: Plataforma para la generación dinámica de aplicaciones distribuidas basadas en la integración de redes de sensores y dispositivos electrónicos IoT,” UNIVERSIDAD DE VIEDO, 2013.
5. L. Atzori, A. Iera, G. Morabito, and M. Nitti, “The Social Internet of Things (SIoT) – When social networks meet the Internet of Things: Concept, architecture and network characterization,” *Comput. Networks*, vol. 56, no. 16, pp. 3594–3608, Nov. 2012.
6. C. González García, B. C. Pelayo G-Bustelo, J. Pascual Espada, and G. Cueva-Fernandez, “Midgar: Generation of heterogeneous objects interconnecting applications. A Domain Specific Language proposal for Internet of Things scenarios,” *Comput. Networks*, vol. 64, pp. 143– 158, May 2014.
7. R. Roman, J. Zhou, and J. Lopez, “On the features and challenges of security and privacy in distributed internet of things,” *Comput. Networks*, vol. 57, no. 10, pp. 2266–2279, Jul. 2013.

8. J. Pascual Espada, O. Sanjuán Martínez, B. C. Pelayo GBustelo, and J. M. Cueva Lovelle, "Virtual Objects on the Internet of Things," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 1, no. 4, p. 23, 2011.
9. B. Xu, L. Da Xu, H. Cai, C. Xie, J. Hu, and F. Bu, "Ubiquitous Data Accessing Method in IoT-based Information System for Emergency Medical Services," *IEEE Trans. Ind. Informatics*, vol. 3203, no. c, pp. 1–1, 2014. 100
10. La Ciberseguridad en la Industria 4.0 <https://www.incibe-cert.es/blog/ciberseguridad-industria-4-0>

Análisis de las Tecnologías 4.0 en las PyMES del partido de Hurlingham

Marisa Panizzi¹, Fernando Puricelli¹, Agustín Hodes¹, Felipe Ortiz¹, Cristian Schiffino¹, Joaquín Pettinari¹,
Florescia Massey¹, Ayelén Rodríguez¹, Rodolfo Bertone².

¹ Instituto de Tecnología e Ingeniería. Universidad Nacional de Hurlingham.
Av. Vergara 2222 (B1688GEZ) - Villa Tesei - Bs. As. Argentina.

² Instituto de Investigación en Informática (III-LIDI). Facultad de Informática.
Calles 50 y 120 - La Plata - Bs. As. - Argentina

marisa.panizzi@unahur.edu.ar, fernando.puricelli@unahur.edu.ar, agustin.hodes@unahur.edu.ar,
felipe.ortiz@unahur.edu.ar, Cristian.schiffino@unahur.edu.ar, joaquin.salvador.pettinari@estudiantes.unahur.edu.ar,
florescia.massey@estudiantes.unahur.edu.ar, ayelen.rodriguez@estudiantes.unahur.edu.ar,
pbertone@lidi.info.unlp.edu.ar

Resumen

Las PyMES requieren incrementar la capacidad y calidad de sus procesos para lograr mayor competitividad dentro de sus sectores industriales. Las soluciones tecnológicas 4.0 enfocadas en la interconectividad, la automatización y los datos en tiempo real contribuyen a tal fin. La adopción de estas tecnologías por parte de las industrias se denomina “industria 4.0¹”. El análisis de la necesidad de las PyMES del partido de Hurlingham de la adopción de este tipo de tecnologías será del presente proyecto de investigación. En la primera etapa de este proyecto, se propone la realización de un estudio exploratorio basado en el método de encuesta. Luego, los datos obtenidos de la situación actual de las tecnologías 4.0 de las industrias en el partido de Hurlingham, se analizarán mediante la aplicación de procesos de minería de datos que permitirá lograr un mapa de la situación actual de la región.

Palabras clave: Tecnologías 4.0, procesos industriales, minería de datos, PyMES, partido de Hurlingham.

Contexto

La línea de investigación que se reporta en este artículo es financiada por un proyecto de investigación titulado “Estudio de la práctica actual de las tecnologías 4.0 en las PyMES del partido de Hurlingham mediante una encuesta” por la Secretaría de Investigación de la

Universidad Nacional de Hurlingham (Resolución Consejo Superior 000382-21).

Esta investigación cuenta con el asesoramiento científico del grupo de investigación de Ingeniería de Software del Instituto de Investigación en Informática (III-LIDI) de la UNLP.

Introducción

En Argentina, los sectores industriales se componen mayoritariamente por Pequeñas y Medianas Empresas (PyMES), lo que constituye un eslabón fundamental para el país y esto refuerza la necesidad de llevar adelante iniciativas que contribuyan con el desarrollo y mejora de competitividad de dichas empresas. En la relación a la industria del software y servicios informáticos, las PyMES representan casi el 80% del sector [1]. Este sector presenta un alto potencial para generar valor agregado al ecosistema productivo del país, promoviendo la generación de empleo calificado y evidenciando un crecimiento exponencial en los últimos años. De hecho, el notable desarrollo queda demostrado en el último informe publicado por el Observatorio de la Economía del Conocimiento (OEC) de Argentina [2]. Esto constituye un eslabón fundamental, en el sector, para el país y refuerza la necesidad de llevar adelante iniciativas que contribuyan con el desarrollo y mejora de competitividad de dichas empresas. A nivel internacional se presenta la misma situación en la industria del software, las

¹ <https://www.argentina.gob.ar/produccion/planargentina40/industria-4-0>

PyMES ocupan una gran porción de la industria en varios países [3]. Se observa que en los últimos años las PyMES han surgido muy rápidamente y, en la mayoría de las economías en desarrollo, el sector está dominado por pequeñas y recientes jóvenes [4].

Las PyMES se han dado cuenta de que es fundamental para su negocio mejorar sus procesos y métodos de trabajo, pero carecen del conocimiento y los recursos para hacerlo. La única forma de contribuir al éxito de los proyectos, por tanto, es definir, implementar y estabilizar los procesos de desarrollo [5].

Para determinar la manera en que las PyMES del partido de Hurlingham pueden fortalecer sus procesos mediante la adopción de tecnologías 4.0 [6] se requiere recolectar evidencia sobre el estado actual de la aplicación de la tecnología en sus procesos productivos. Para lograr esto, nuestro primer estadio de la investigación consiste en realizar un estudio exploratorio a través de una encuesta. Para la realización de la encuesta, específicamente el relacionamiento con las PyMES de la región se trabajará de manera colaborativa con el Centro PyME-UNAHUR [7].

Los hallazgos obtenidos mediante el estudio son de utilidad para el grupo de investigación porque proporcionan una caracterización de la aplicación de las TICS en las PyMES participantes de la encuesta. A partir del análisis de los datos obtenidos mediante la aplicación de minería de datos, se podrán identificar las necesidades respecto a la adopción de tecnologías 4.0.

Líneas de Investigación, Desarrollo e Innovación

El objetivo general de esta línea de investigación consiste en determinar el estado actual de las PyMES del partido de

Hurlingham respecto a la aplicación de las tecnologías 4.0 en sus procesos industriales.

Los hallazgos obtenidos mediante la encuesta permitirán identificar las necesidades de las PyMES respecto a 4 áreas de conocimiento vinculadas a las tecnologías 4.0. Estas áreas de conocimiento abarcan el desarrollo de software, la ciberseguridad, la ciencia de datos y la inteligencia artificial desde el punto de vista técnico. Además, se considerará en el análisis una perspectiva social respecto al grado de consideración de la ley de género relacionada a la equidad de la distribución de los puestos de trabajo incluyendo los cargos de conducción por parte de estas empresas.

Resultados y Objetivos

Este proyecto de investigación se inicia en este año por lo tanto no cuenta con resultados alcanzados. A continuación, se describen los objetivos que se pretenden lograr:

- a) Académicos, dos tesinas de grado de la carrera Licenciatura en Informática.
- b) Producción Científica: se presentarán avances de la investigación en eventos científicos de alcance nacional (WICC² 2023, CACIC³ 2022 y CACIC 2023) y en el ámbito internacional, CIACA 2022⁴ e InGENIO⁵ 2022
- c) Formación en investigación: el grupo de investigación inicia un proceso de aprendizaje de métodos de investigación de ingeniería de software experimental, revisiones sistemáticas [8] y encuestas [9][10].

Para el desarrollo de este proyecto de investigación, se seguirá un enfoque de investigación clásico [11],[12]. Los métodos y materiales necesarios para desarrollo de la primera etapa del proyecto son los siguientes:

- Métodos.
Mapeo Sistemático de la literatura (en inglés *Systematic Mapping Study* o SMS). Para la revisión sistemática de la literatura se empleará

² Workshop de Investigadores en Ciencias de la Computación

³ Congreso Argentino de Ciencias de la Computación.

⁴ Conferencia Iberoamericana de Computación Aplicada.

⁵ Congreso Latinoamericano de Ingeniería

un método de investigación secundario de acuerdo con las directrices propuestas en [13],[14]. Un método de investigación secundario permite recopilar de manera sistemática y rigurosa los estudios primarios relacionados con una pregunta de investigación específica, con el objetivo de sintetizar la evidencia disponible para responder dicha pregunta. En nuestro proyecto nos interesa conocer si existen otros estudios similares respecto a la determinación de características de empresas, como por ejemplo la encuesta HELENA (*Hybrid DEveLopmENt Approaches in software systems development*) [15] a nivel internacional o la encuesta realizada por la CESSI (Cámara de la Industria Argentina de Software) a nivel nacional [1].

Las encuestas son investigaciones que proporcionan una visión general, mediante la recolección de información estandarizada de una población específica o una muestra representativa de la misma (sujetos del estudio), por medio de un cuestionario o entrevista [9]. En la mayoría de los casos, los datos relativos a la encuesta provendrán de cuestionarios. Pero los cuestionarios, por sí solos no constituyen la encuesta. Para la construcción del cuestionario, se emplearán las directrices propuestas en [8]. Las encuestas son métodos de investigación primarios al igual que los experimentos, estudio de casos y entrevistas; permiten obtener evidencia empírica sobre algún tema de interés. De hecho, una encuesta [16] es un proceso más complejo formado por una serie de actividades bien definidas que se enumeran a continuación: a) establecer los objetivos de la encuesta, b) diseñar la encuesta, c) desarrollar el cuestionario, c) evaluar y validar el cuestionario, c) obtener los resultados de la encuesta y c) analizar los resultados obtenidos. Prototipado Evolutivo Experimental (Método de la Ingeniería). El prototipado evolutivo experimental [17] consiste en desarrollar una solución inicial para un determinado problema, generando su refinamiento de manera evolutiva por prueba de aplicación de dicha solución a casos de estudio (problemáticas) de complejidad creciente. El proceso de

refinamiento concluye al estabilizarse el prototipo en evolución.

- **Materiales.**

La automatización del instrumento de recolección de los datos de la encuesta se realizará a través del uso de alguna herramienta libre. Para la búsqueda de literatura existente se empleará el Sistema Nacional de Repositorios Digitales (SNRD) [18]. Se utilizará la base de contactos de las PyMES del partido de Hurlingham del Centro PyME-UNAHUR.

Formación de Recursos Humanos

El grupo se encuentra conformado por un Director, un Codirector, tres docentes-investigadores, tres alumnos de grado y un asesor científico-tecnológico.

Se estima la formación de dos tesinas de grado de la carrera Licenciatura en Informática de UNAHUR.

Referencias

[1] OPSSI. Reporte anual 2018 (2018). sobre el Sector de Software y Servicios Informáticos de la República Argentina. Disponible en <https://www.cessi.org.ar/opssi>.

[2] Informe Argentina Productiva – Economía del Conocimiento (2019). Ministerio de Producción y Trabajo. Presidencia de la Nación. Secretaría de la Transformación Productiva. Diciembre de 2019. Disponible en:

<https://biblioteca.produccion.gob.ar/buscar/?fid=16>.

[3] Hisham M. Abushama (2016). PAM-SMEs: process assessment method for small to medium enterprises. *Software: Evolution and Process*, 28, pp. 689 –711 (2016).

[4] Sharma P., Sangal A.L (2019). Building a hierarchical structure model of enablers that affect software process improvement in software SMEs-A mixed method approach. *Computer Standards & Interfaces*, 66, pp. 1–23.

- [5] Ianzen A., Mauda E.C., Paludo M.A., Reinehr S., Malucelli (2013). A. Software process improvement in a financial organization: an action research approach. *Computer Standard & Interfaces*, 36, pp 54–65 (2013).
- [6] Ministerio de Desarrollo Productivo (2021). Plan de Desarrollo Productivo Argentina 4.0. Disponible en: <https://www.argentina.gob.ar/produccion/plan-argentina40/industria-4-0>.
- [7] Centro PyME-Unahur. Vinculación Tecnológica de la Universidad Nacional de Hurlingham. Disponible: <https://unahur.edu.ar/centro-pyme-unahur/>
- [8] Kitchenham, B. y Chartes, S. (2007). Guidelines for performing systematic literature reviews in Software engineering, Keele University, EBSE-2007-01.
- [9] Jefferson Seide Molléri, Kai Petersen, Emilia Mendes (2020). An empirically evaluated checklist for surveys in software engineering. *Information and Software Technology* 119, 106240.
- [10] Genero, M., Piattini, M., & Cruz Lemus, J. A. (2014). *Métodos de investigación en Ingeniería del Software*. Madrid: Ra-Ma S.A. Editorial y Publicaciones.
- [11] Riveros, H. y Rosas, L. (1985). *El Método Científico Aplicado a las Ciencias Experimentales*. Editorial Trillas. México. ISBN 96-8243-893-4. 2.
- [12] Creswell, J. (2002). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Prentice Hall. ISBN 10: 01-3613-550-1. 3.
- [13] Kitchenham, B, Budgen, D., Brereton, P. (2016). *Evidence-Based Software Engineering and Systematic Reviews*. CRC Press.
- [14] Petersen K., Feldt R., Mujtaba S., Mattsson M. (2008). Systematic mapping studies in software engineering, In: *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering*, pp. 68–77.
- [15] Marco Kuhrmann, Paolo Tell, Jil Klünder, Regina Hebig, Sherlock Licorish, Stephen MacDonell (Eds.): *Complementing Materials for the HELENA Study (Stage 2)*. [online] DOI: 10.13140/RG.2.2.11032.65288, published: 2018-11-28.
- [16] Kitchenman, B. A.y Pfleeger, S. L. (2008). Personal opinion survey. Capítulo 3 del Libro *Guide to Advanced Empirical Software Engineering*. Shull F., Janice Singer J., Sjøberg D.I.K. Editors. Springer.
- [17] Basili. *The Experimental Paradigm in Software Engineering*. En *Experimental Software Engineering Issues: Critical Assessment and Future Directions* (Ed. Rombach, H., Basili, V., Selby, R.). *Lecture Notes in Computer Science*, Vol. 706. (1993). ISBN 978-3-540-57092-9.
- [18] Ministerio de Ciencia, Tecnología e Innovación. *Sistema Nacional de Repositorios Digitales*. Disponible en: <https://repositoriosdigitales.mincyt.gob.ar/vufind/>

MODELOS BASADOS EN INTELIGENCIA ARTIFICIAL Y COMPUTACIÓN UBICUA PARA LA RESOLUCIÓN DE PROBLEMAS EN EDUCACIÓN Y OTROS DOMINIOS

Durán Elena, Álvarez Margarita, Únzaga Silvina, Salazar Nevelin, Fernández Reuter Beatriz, Lara Cecilia, González Gabriela, Espeche Fabián, Acosta Denis, Díaz Fátima y Juárez Gastón

Instituto de Investigaciones en Informática y Sistemas de Información
Departamento de Informática - Facultad de Ciencias Exactas y Tecnologías
Universidad Nacional de Santiago del Estero (UNSE)
e-mail: {eduran, alvarez, sunzaga, nsalazar, bfreuter, clara@unse.edu.ar, ggonzalez@unse.edu.ar}@unse.edu.ar, h_espeche@hotmail.com, denislionelacosta@gmail.com, fatimadiaz91@gmail.com, lucianojuarezgaston@gmail.com

CONTEXTO

En este trabajo se presenta el proyecto de investigación “Modelos basados en Inteligencia Artificial y Computación Ubicua para la resolución de Problemas en Educación y otros dominios”, correspondiente a la convocatoria 2021 de la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Santiago del Estero (SICYT - UNSE). El proyecto tiene un período de ejecución desde el año 2022 hasta el año 2025. Lo allí propuesto es una continuación de las líneas de investigación iniciada en 2012, en el proyecto "Sistemas de información web personalizados, basados en ontologías, para soporte al aprendizaje ubicuo" y continuadas en el 2017 en el proyecto “Métodos y Técnicas para desarrollos de Aplicaciones Ubicuas”.

El proyecto de investigación presentado en este trabajo, propone favorecer el desarrollo de conocimiento científico- tecnológico de relevancia, principalmente sobre Computación Ubicua e Inteligencia Artificial, realizando propuestas de modelos computacionales para la resolución de problemas en Educación y otros dominios.

RESUMEN

Dos nuevas tecnologías han irrumpido fuertemente en el mundo en los últimos años, la Inteligencia Artificial (IA) y la Computación Ubicua (CU). La primera de ellas se caracteriza por crear máquinas inteligentes que simulen el razonamiento y comportamiento humanos. Por su parte, la CU hace referencia a una tecnología profunda que está inmersa en la vida cotidiana, de forma tal que no se

puede distinguir de ella. La CU se aplica en diversos campos, siendo el campo de la educación una rama destacada en las investigaciones. Para hacer posible el aprendizaje de los estudiantes, los modelos computacionales requieren aplicar técnicas de IA para adaptar el contenido a cada alumno de acuerdo al contexto, a los perfiles de aprendizaje y a las características de los dispositivos que usan para concretar el aprendizaje.

En consecuencia, la combinación de la IA y la CU puede generar resultados de alto impacto en educación y en otros dominios. Es por eso que, en este proyecto, se quiere abordar el diseño de nuevos modelos de IA y CU para atender problemáticas relacionadas principalmente con el desarrollo y la evaluación de estrategias y recursos de aprendizaje. Se espera además, favorecer el desarrollo de conocimiento científico-tecnológico de relevancia y formar recursos humanos en las áreas de CU e IA.

Palabras clave: Modelos Computacionales, Inteligencia Artificial, Computación Ubicua, Realidad Aumentada, Procesos de Enseñanza y de Aprendizaje.

1. INTRODUCCIÓN

Los usuarios necesitan sistemas con una operación simple y fácil de recordar; dan cada vez más importancia a los sistemas que brindan ayuda de forma automática para completar las tareas. En este sentido, el nuevo paradigma de la CU permite ofrecer entornos con sistemas donde los usuarios pueden obtener automáticamente información y funcionalidad según su estado (por ejemplo:

ubicación, tarea actual, etc.). Este paradigma incorpora la capacidad computacional en objetos cotidianos para que se comuniquen de manera efectiva y realicen tareas útiles (Chaparro-González, 2003).

En los entornos ubicuos los ordenadores están incluidos en nuestros movimientos naturales y en las interacciones con nuestro entorno, tanto físico como social. La CU involucra la movilidad; es decir, la capacidad de mover los servicios informáticos con nosotros; y la pervasividad, o sea, la capacidad de obtener información del entorno en el que está inmersa una aplicación informática y utilizarla para construir dinámicamente modelos computacionales (Hwang et al., 2008), (Sakamura y Koshizuka, 2005).

Muchos campos y dominios de aplicación se benefician de la CU, principalmente tiene un gran potencial en el campo de la educación. Así ha surgido un nuevo paradigma educativo denominado Aprendizaje Ubicuo que permite el aprendizaje del contenido correcto, en el lugar más apropiado, en el momento indicado y de la manera correcta Yahya *et al.* (2010). Para hacer posible este aprendizaje, los modelos computacionales requieren aplicar técnicas de IA que permitan adaptar el contenido a cada alumno de acuerdo al contexto, a los perfiles de aprendizaje y a las características de los dispositivos que usan para concretar el aprendizaje.

Por lo tanto, el desarrollo de aplicaciones de software ubicuas presenta para los desarrolladores un inmenso desafío.

En proyectos anteriores se ha dado inicio a esta línea de investigación, ahondando en el estudio de los métodos y técnicas para desarrollar aplicaciones ubicuas (Álvarez et al., 2015), (Álvarez et al., 2017), (Durán et al., 2016), (Durán y Álvarez, 2017), (Fernández Reuter et al., 2017), (González y Durán, 2014) y (Únzaga et al., 2015). No obstante, son muchos los interrogantes que aún persisten; como, por ejemplo: ¿cómo modelar contexto heterogéneo y dinámico de forma eficiente?, ¿cómo proporcionar modelos de adaptación al contexto para usuarios particulares?, ¿cómo crear modelos basados en IA para automatizar estrategias y recursos de aprendizaje? ¿cómo aplicar y evaluar los modelos en desarrollos de software educativo y de otros dominios? Es decir, el desarrollo de un sistema ubicuo requiere de la utilización de técnicas especializadas de adquisición y modelado del contexto, técnicas de IA, tales como: razonamiento automático,

adaptación, agentes inteligentes, aprendizaje automático, ontologías, planning, etc.

En este artículo se presentan los objetivos, la metodología, los resultados esperados y la formación de recursos humanos del proyecto de investigación.

2. OBJETIVOS

Con este proyecto de investigación se busca realizar contribuciones teóricas y metodológicas en el campo de los Modelos basados en IA y CU en el ámbito del aprendizaje y otros dominios de aplicación. En consecuencia, los objetivos son:

1. Desarrollar el Estado del Arte del diseño y construcción de Modelos basados en IA y CU en el ámbito de la Educación y otros dominios.
2. Diseñar modelos basados en IA y/o CU para la generación y evaluación de estrategias y recursos de aprendizaje.
3. Diseñar un modelo del contexto ambiental basado en el descubrimiento automático de patrones en sistemas ubicuos para diferentes dominios.
4. Aplicar los modelos propuestos en la construcción de aplicaciones de software para la resolución de problemas en educación y otros dominios.
5. Evaluar, en contextos reales, el nivel de satisfacción del usuario y el desempeño de las aplicaciones construidas en base a los modelos diseñados.

3. METODOLOGÍA

Con el propósito de dar cumplimiento al objetivo específico 1, se realizarán las siguientes actividades:

- a. Exploración e investigación bibliográfica sobre modelado computacional con IA y/o CU de problemas de educación y otros dominios.
- b. Análisis de los antecedentes encontrados.
- c. Síntesis de los antecedentes encontrados.

Con el fin de obtener el objetivo específico 2, se realizará:

- a. Diseñar un modelo basado en IA para la evaluación de la Calidad de los Objetos de Aprendizaje desde el enfoque de los estudiantes.
- b. Diseñar un modelo para la Generación automática de caminos de aprendizajes personalizados para aprendizaje ubicuo.
- c. Diseñar un modelo para los procesos de evaluación en el aprendizaje ubicuo.

- d. Diseñar un modelo para aprendizaje ubicuo con el uso de Laboratorios Virtuales.

Para el objetivo específico 3 se plantean las siguientes actividades:

- a. Diseño del modelo.
- b. Diseño de la estrategia de adquisición del contexto en tiempo real.
- c. Inferencia de eventos y patrones de eventos.
- d. Evaluación del modelo de contexto construido.

Con el propósito de dar cumplimiento al objetivo específico 4, se realizarán las siguientes actividades:

- a. Identificación de problemáticas educativas y de otros dominios que justifiquen el desarrollo de aplicaciones ubicuas con los modelos computacionales previamente diseñados .
- b. Diseño de las aplicaciones ubicuas.
- c. Construcción del software.

Por último, para dar cumplimiento con el objetivo específico 5 se realizará:

- a. Operacionalización de las variables nivel de satisfacción del usuario y desempeño de aplicaciones ubicuas.
- b. Ejecución de las aplicaciones en contextos reales.
- c. Análisis de resultados.
- d. Elaboración de conclusiones.

4. RESULTADOS ESPERADOS

Los resultados que se esperan obtener de esta investigación son:

- Estado del arte sobre modelos computacionales basados en IA y CU para educación y otros dominios
- Modelos computacionales de contexto ambiental para sistemas ubicuos en diferentes dominios.
- Modelos computacionales para la Generación automática de caminos de aprendizajes personalizados para aprendizaje ubicuo
- Modelo de evaluación de la Calidad de Objetos de Aprendizaje desde el enfoque de los estudiantes.
- Modelo para aprendizaje ubicuo con el uso de Laboratorios Virtuales.
- Modelo para los procesos de evaluación en el aprendizaje ubicuo.
- Modelo de innovación educativa para el aprendizaje de la programación en entornos de aprendizaje ubicuo.

- Prototipo que implemente el Modelo de contexto ambiental para sistemas ubicuos.
- Prototipo que implemente el Modelo para la Generación automática de caminos de aprendizajes personalizados para aprendizaje ubicuo.
- Prototipo que implemente el Modelo para los procesos de evaluación en el aprendizaje ubicuo.
- Prototipo de aplicación que recomiende en forma personalizada Tutores para entornos aprendizaje ubicuo.
- Prototipo de aplicación para el Aprendizaje Ubicuo de Conceptos Básicos de Programación con Realidad Aumentada.
- Prototipo de sistemas de recomendación ubicuo de documentos científicos para Repositorios Digitales Institucionales

El proyecto tendrá también un importante impacto a nivel local, ya que los desarrollos concretados en el marco del proyecto serán transferibles de modo directo a organizaciones del medio cuyas problemáticas se atienden en estos desarrollos.

5. FORMACIÓN DE RECURSOS HUMANOS

El desarrollo del proyecto de investigación presentado, facilitará la formación de recursos humanos de la Facultad de Ciencias Exactas y Tecnologías de la UNSE. En este sentido, dos integrantes están desarrollando su Tesis de Doctorado en Ciencias de la Computación. Tres integrantes del proyecto, elaborarán su Tesis de la carrera de Maestría en Informática Educativa de la UNSE. Por último, tres alumnos de la Licenciatura en Sistemas de Información llevarán a cabo sus tesis de grado.

Además, con el desarrollo de este proyecto se está afianzando el grupo de investigación en las temáticas del proyecto: Modelos Computacionales, IA, CU, Realidad Aumentada y Procesos de Enseñanza y de Aprendizaje, lo que contribuye a una mejora en el fondo de conocimiento disciplinar disponible, no sólo a nivel local sino también regional y nacional.

6. BIBLIOGRAFÍA

- Álvarez, M., Únzaga S. y Durán E. (2015). "Modelo de dominio en sistemas de aprendizaje ubicuo". 10° Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA. 21 y 22

- de mayo de 2015. Universidad Nacional de Salta. ISBN N°: 978-987-633-133-3. Pág. 116.
- Álvarez, M., Únzaga S. y Durán E. (2017). "Método para generar recomendaciones personalizadas para integrar grupos de aprendizaje ubicuo y colaborativo". XXIII Congreso Argentino de Ciencias de la Computación (CACIC 2017). Libro de actas. Pag. 363- 372. ISBN 978-950-34-1539-9. Compiladores: De Giusti, Armando Eduardo y Pesado, Patricia Mabel. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/63019>. La Plata. Prov. Buenos Aires. 9 al 13 de octubre.
- Chaparro-González, D. (2003) Computación ubicua. (Tesina de grado, Universidad Rey Juan Carlos, Madrid, España). Disponible en <http://docplayer.es/6315315-Universidad-reyjuan-carlos>.
- Durán, E., Álvarez, M. y Únzaga S. (2016). "Modelo ontológico para personalizar aplicaciones de aprendizaje ubicuo". EATIS 2016. Colombia. 27-29 de abril de 2016.
- Durán, E. y Álvarez, M. (2017). "Método para generar recomendaciones de expertos para asesorar a los estudiantes sobre experiencias de aprendizaje ubicuas", 36° Congreso Internacional de la Sociedad Chilena de Ciencias de la Computación (SCCC), 2017, pp. 1-8, doi: 10.1109 / SCCC.2017.8405121
- Fernández Reuter, B., Durán, B. y Amandi, A (2017). "Designing a Hybrid Method for Personalized Ubiquitous Learning Paths Generation". Conferencia Internacional de la Sociedad Chilena de Ciencia de la Computación (SCCC 2017). Arica, Chile. 16 al 20 de Octubre de 2017.
- González, G. y Durán E. (2014). "Modelo del estudiante para sistemas de aprendizaje ubicuo: representación por medio de ontologías". IX Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET 2014), Universidad Nacional de Chilecito, La Rioja, ISBN 978-987-24611-1-9, pp. 298-305.
- Hwang, G.J., Tsai, C.-C. y Yang, S. J. H. (2008). "Criteria, strategies and research issues of context-aware ubiquitous learning". Journal of Educational Technology & Society, vol. 11(2), pp. 81-91.
- Sakamura, K. y Koshizuka, N. (2005). "Ubiquitous computing technologies for ubiquitous learning". IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE'05), Tokushima, Japan.
- Únzaga S., Álvarez M., Durán E. (2015). "Modelo de Requerimientos de una Aplicación de Apoyo al Aprendizaje Ubicuo para el Ingreso Universitario". TE&ET'15: X Congreso de Tecnología en Educación y Educación en Tecnología 2015. Argentina
- Yahya, S., Ahmad, E., & Abd Jalil, K. (2010). The definition and characteristics of ubiquitous learning: A discussion. International Journal of Education and Development using ICT, 6(1).

Adecuación de un Sistema de Recuperación de Información para su utilización en un Contexto Jurídico

Oswaldo Sposito¹, Hugo Ryckeboer¹, Julio Bossero¹, Edgardo Moreno¹, Viviana Ledesma¹, Gastón Procopio¹, Lorena Matteo¹, Cecilia Gargano¹, Victoria Saizar¹, Patricio Macias¹, Juan Ojeda¹, Fabio Quintana¹, Laura Conti², Sergio García³ y Gustavo Pérez Villar⁴

¹ Universidad Nacional de La Matanza. Departamento de Ingeniería e Investigación Tecnológicas. Florencio Varela 1903. San Justo. La Matanza.

{sposito, hugor, jbossero, ej_moreno, vledesma, gprocopio, lmatteo, cgargano, vsaizar, pmacias, fquintana, jmojeda}@unlam.edu.ar

² Universidad Nacional de La Matanza. Departamento Derecho y Ciencia Política. lconti@unlam.edu.ar

³ Palacio de Tribunales. Departamento Judicial de Morón. Alte. Brown. Piso 4. Morón. sergiogabriel.garcia@pjba.gov.ar

⁴ Subsecretaría de Tecnología Informática del Poder Judicial de la Provincia de Buenos Aires. Palacio de Justicia, avenida 13 entre 47 y 48, primer piso (La Plata). Argentina. gperez@scba.gov.ar

RESUMEN

En las últimas décadas, las instituciones públicas, particularmente el Poder Judicial (PJ), con el desarrollo de las TICs, han generado un importante aumento en: la generación de documentos digitales, en los repositorios de los mismos y en los Sistemas de Recuperación de Información (SRI). Este trabajo se orienta a estudiar y proponer soluciones para la recuperación de documentos judiciales, se hace una propuesta para la construcción de la matriz de términos en un proceso de indización.

Palabras clave: SRI, Modelo Vectorial, Indización, Lematización.

CONTEXTO

La línea de investigación aquí presentada es parte del proyecto de investigación “Implementación de un Sistema Web de Recuperación de la Información Orientado a Documentación Jurídica con el Proceso de Indexación Semántica Latente Paralelizado”, perteneciente al programa de Investigaciones PROINCE (Programa de Incentivos para Docentes Investigadores) de la Secretaría de Políticas Universitarias del Ministerio de Educación de la Nación. Los integrantes del equipo son docentes e investigadores

dependientes de las siguiente Unidades Académicas de la Universidad Nacional de La Matanza (UNLaM): el Departamento de Ingeniería e Investigaciones Tecnológicas (DIIT), y el Departamento Derecho y Ciencia Política, además, colaboran personal técnico de la Subsecretaría de Tecnología Informática del Poder Judicial de la Provincia de Buenos Aires.

1. INTRODUCCIÓN

En este trabajo se describe un proceso de indización, que consiste en extraer una serie de términos, representativos de los temas tratados en un documento, para utilizarlos después como puntos de acceso para la recuperación de esos documentos de un corpus jurídico. El propósito es brindar jurisprudencia similar a los profesionales del derecho luego de realizar una consulta. Entendiendo el concepto de jurisprudencia, como el conjunto de las sentencias de distintos fallos dictados por los tribunales de justicia u organismos judiciales de un Estado. En el campo del derecho, la jurisprudencia juega un papel importante como fuente del derecho; por ser la comprensión e interpretación de las normas jurídicas basada en las sentencias pasadas emitidas por órganos oficiales,

estas sustentan la aplicación de la ley en un caso concreto. En el PJ se producen una enorme cantidad de documentos jurídicos (dictámenes, expedientes, etc.) cada año, lo cual produce que esta fuente de derecho sea cada vez mayor, lo que impulsa a los profesionales del derecho a dedicar más tiempo a la búsqueda de una decisión relevante.

Basándonos en [1] coincidimos, en que los SRI están en continua mejoría, esto se debe a: la incorporación de utilidades dependientes de la expansión de su uso, el avance de las aplicaciones tecnológicas y el claro deslinde de sus funciones.

En [2], se referencia a Calvin N. Mooers como quien introdujo por primera vez en 1950 el término Recuperación de Información (en inglés Information Retrieval) en la literatura de documentación, la definió como «*la búsqueda de información en un stock de documentos, efectuada a partir de la especificación de un tema*». Sólo un año más tarde, el mismo autor ampliaba esta definición al manifestar que la recuperación de información abarca los aspectos intelectuales de la descripción de información y su especificación para la búsqueda, y también cualquier sistema, técnica o máquina que se utilice para llevar a cabo la operación [3].

Según la bibliografía consultada [4-6], una SRI es un programa que interactúa entre un corpus y sus usuarios. Su efectividad depende del adecuado control del lenguaje de representación de los elementos de información y las búsquedas de sus usuarios. Para cumplir con sus objetivos, según Gabriel H. Tolosa y otros [5], un SRI debe realizar las siguientes tareas básicas:

- Representación lógica de los documentos y, opcionalmente, almacenamiento del original.
- Representación de la necesidad de información del usuario en forma de consulta.

- Evaluación de los documentos respecto de una consulta para establecer la relevancia de cada uno.
- Ranking de los documentos considerados relevantes para formar el “conjunto solución o respuesta. Presentación de la respuesta al usuario.
- Retroalimentación de las consultas para aumentar la calidad de la respuesta.

Jaime Robredo en [5], asevera que en cualquier área del conocimiento, los términos con significado se pueden utilizar como descriptores para representar el contenido de documentos escritos, en los procesos de indización y organización de la información, así como para formular preguntas en el proceso de recuperación de información. Tolosa en [5] afirma que el proceso se puede dividir en las siguientes etapas:

- Análisis lexicográfico: Se extraen las palabras y se normalizan.
- Reducción (Tokenización) de palabras vacías o de alta frecuencia.
- Lematización: Se reducen palabras morfológicamente parecidas a una forma base o raíz, con la finalidad de aumentar la eficiencia de un SRI.
- Asignación de pesos o ponderación de los términos que componen los índices de cada documento.

Los SRI implementan una gama diversa de estructuras de datos, algoritmos y técnicas de recuperación de información, por ello, se precisa de un modelo conceptual donde se determinen: el tipo de almacenamiento, operaciones sobre los términos, modelos de búsqueda con base patrones exactos o los modelos inexactos los cuales contendrán las técnicas probabilísticas, los modelos lógicos y los espacios vectoriales [8]. En el trabajo de Martínez Méndez, se puede encontrar un estudio detallado de los distintos modelos de RI existentes. Uno de los modelos más utilizados [4-5], es el Modelo de Espacio

Vectorial. En este modelo, el texto es representado por un vector de términos, los términos comúnmente son palabras; cualquier texto puede ser representado por un vector en un espacio dimensional Salton en el año 1975 [9]. En un espacio de documento que consiste en documentos D_i , cada uno identificado por uno o más términos de índice T_j ; los términos pueden ser ponderados de acuerdo a su importancia, o no ponderados con pesos restringidos a 0 y 1. En el modelo los documentos se representan a partir de vectores, de la siguiente manera:

$$D_i = (T_1, T_2, \dots, T_j) \quad (1)$$

En la Figura 1 se muestra un espacio de índice tridimensional, donde cada elemento se identifica con hasta tres términos distintos.

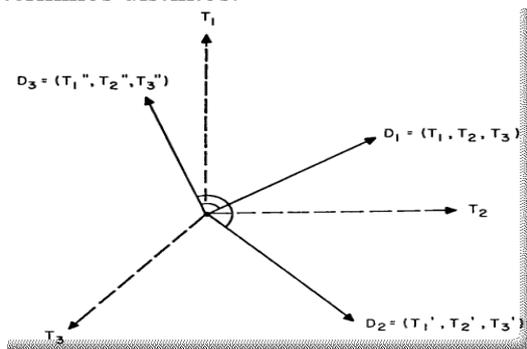


Figura. 1. Representación vectorial espacial de los documentos [9].

Una consulta se puede ver como un documento por lo tanto se puede ver como un vector.

Matemáticamente, una de formas de establecer la cercanía de dos vectores es calcular el coseno del ángulo que forman los dos vectores entre sí. Esta fórmula tiene la ventaja de su bajo esfuerzo computacional y es independiente de los módulos de los vectores. De manera similar, se puede calcular el coseno del

ángulo entre cada vector de documento y el vector de consulta para encontrar su cercanía. Para encontrar un documento relevante para el término de la consulta, se calcula la puntuación de similitud entre cada vector del documento y el vector del término de la consulta aplicando la similitud del coseno. Finalmente, aquellos documentos con puntajes de similitud altos se considerarán documentos relevantes para la consulta. [9].

Como se comentó, dentro de la indización se encuentra la lematización, que es una técnica empleada en la recuperación de datos en los SRI, que sirve para reducir variantes morfológicas de la forma de una palabra a raíces comunes o lexemas; con el fin de mejorar la habilidad de los motores de búsqueda y, a consecuencia, los resultados de las consultas. Básicamente, este consiste en remover el plural, el tiempo, o los atributos finales de la palabra [5,6,10]. En el trabajo de González [6], afirma que “cuando se realiza la extracción de palabras de un texto se obtiene una gran cantidad de entradas con formas verbales conjugadas y variantes de concordancia. Logrando la reducción morfológica de todas estas variantes se busca que el usuario recupere tanto los textos que contienen sus términos de búsqueda, como aquellos que contienen las formas derivadas de esos términos...”. Cabe aclarar que, en este proyecto, nosotros también simplificamos las apariciones de sustantivos y adjetivos. Los algoritmos de lematización más conocidos son: Lovins¹ (1968), Porter² (1980) y Paice³ (1990). Originalmente todos fueron hechos para el inglés, y se diferencian en la eficiencia

1

<http://snowball.tartarus.org/algorithms/lovins/stemmer.html>

² <https://tartarus.org/martin/PorterStemmer/>

³ <https://www.scientificpsychic.com/paice/paice.html>

del código y la elección de sufijos que identifican y eliminan. Una modificación del algoritmo trabajo de Porter, es el algoritmo de Snowball⁴. Este puede mapear palabras que no están en inglés. Estos algoritmos permiten realizar “derivaciones”, esto es remover los sufijos comunes morfológicos e inflexionales de palabras literalmente diferentes, pero con una “raíz” común, que pueden ser consideradas como un sólo término. Este algoritmo requiere de un conjunto de pasos para llegar a la raíz.

2. LÍNEAS de INVESTIGACIÓN y DESARROLLO

El presente trabajo tiene como eje central el desarrollo de un SRI. Entre las líneas de investigación a considerar en este proyecto se pueden mencionar:

- El problema de la recuperación de información, el modelo vectorial y la forma de almacenar los términos de una colección (corpus) de pruebas.
- La paralelización del proceso de Indexación Semántica Latente (ISL). Se estudian las librerías: Compute Unified Device Architecture (CUDA) y CUDA Basic Linear Algebra Subprograms (CuBlas), aplicadas a una arquitectura híbrida.
- La aplicando el patrón de arquitectura Modelo-Vista-Controlador (MVC), para desarrollos WEB. Aplicando el lenguaje de programación C#.
- Estudio de la librería REGEX., para resolver las Expresiones Regulares (ER).
- Estudio y evaluación de distintos algoritmos de ranking para Documentos. Las pruebas serán realizadas tomando como base un corpus jurídico real.

3. RESULTADOS OBTENIDOS/ESPERADOS

Durante el año 2021 se ha trabajado, principalmente, en dos temas, por un lado, en el estudio, análisis y modificación de algoritmos y técnicas que permitan la lematización de términos, y por otro, en el proceso que permita incorporar, de un corpus jurídico, las fechas y las referencias de la norma jurídica actual, mediante el Reconocimiento de Entidades Nombradas (tales como Acordadas, Artículos, Leyes, entre otros), que componen los distintos textos judiciales, utilizando Expresiones Regulares (ER). Se presentaron en distintos congresos las siguientes publicaciones:

1. *“Propuesta para la construcción de un corpus jurídico utilizando Expresiones Regulares”*. Presentado en el XXVII Congreso Argentino de Ciencias de la Computación (CACIC). Salta. Argentina [8].

Una ER es una notación algebraica para caracterizar un conjunto de cadenas [11]. Son particularmente útiles para la búsqueda en textos, cuando se tiene un patrón y un corpus de textos donde buscar. En este trabajo se demostró que es posible incorporar en el proceso de Análisis lexicográfico Expresiones Regulares para incorporar fechas y Entidades Nombradas a una matriz de términos. Dentro de las tareas a desarrollar, durante este año, se puede mencionar:

- Incorporar la codificación propuesta al SRI implementado por el proyecto PROINCE mencionado en la introducción.
 - Analizar otros algoritmos y técnicas de derivación.
 - Estudiar otras librerías existentes de ER.
 - Realizar una clasificación de todas las EN dentro de la norma jurídica Argentina.
2. *“Implementación de un lematizador*

⁴ <https://snowballstem.org/demo.html>

para la lengua española". Trabajo presentado en el Workshop del IX Congreso Nacional de Ingeniería en Informática/Sistemas de Información. CONAISI 2021. Mendoza. Argentina. En este trabajo se muestra una modificación realizada al algoritmo de Snowball. Mejorando en un 26% la lematización de términos. Se prevé para este año:

- Modificar el orden de los pasos, propuesto en el algoritmo de Snowball, para mejorar los tiempos de procesamiento.
- Estudiar nuevos métodos de derivación.
- Profundizar en el estudio de la morfología léxica, ciencia que estudia la estructura de las palabras y las pautas que permiten formarlas o derivarlas a partir de otras.

4. FORMACIÓN DE RECURSOS HUMANOS

La presente línea de investigación la lleva adelante un equipo de 15 integrantes provenientes de dos departamentos de la UNLaM, el DIIT y el Departamento de Derecho y Ciencia Política.

- 1 alumno de grado. En el año 2021 se graduó en la carrera de Ingeniería de Informática.
- 2 asesores especialistas externos. (uno perteneciente al Poder Judicial de la Provincia de Buenos Aires y un Secretario de Juzgado).

5. BIBLIOGRAFÍA

- [1] Galindo Ayuda, F. (2020). Avances en sistemas jurídicos de recuperación de documentos. *Scire: Representación Y organización Del Conocimiento*, 26(1), 63–74. <https://doi.org/10.54886/scire.v26i1.4698>. Fecha de consulta: 07/02/22
- [2] S. Oliván, J.A., & Arquero Avilés, Rosario. (2006). Una aproximación al concepto de recuperación de información en el marco de la ciencia de la documentación. *Investigación bibliotecológica*, 20(41), 13-43. Disponible en: <http://www.scielo.org.mx/scielo.php?script=s> ci_arttext&pid=S0187-358X2006000200002 &lng=es&tlng=es. F. de consulta: 07/02/22
- [3] C.N. Mooers, "The theory of digital handling of non-numerical information and its implications to machine economics", en *Technical Bulletin No. 48*. Cambridge, MA: Zator Co., 1950 (Ponencia presentada en Association for Computing Machinery, Rutgers Univ., New Brunswick, NJ, 1950, March 29).
- [4] Kuna, H., Rey, M., Martini, E., Solonezen, L. & Podkowa, L. Desarrollo de un Sistema de Recuperación de Información para Publicaciones Científicas del Área de Ciencias de la Computación. *Rev. Latinoamericana de Ingeniería de Software*, (2014). 2(2): 107-114. <http://revistas.unla.edu.ar/software/article/view/81>. Fecha de consulta: 07/02/22
- [5] Tolosa G. & Bordignon, F. Introducción a la Recuperación de Información: Conceptos, modelos y algoritmos básicos. UNDeL, Argentina, (2008). En línea: <http://eprints.rclis.org/12243/1/Introduccion-RI-v9f.pdf>. Fecha de consulta: 07/02/22
- [6] González, C. M. La recuperación de información en el siglo XX. Revisión y aplicación de aspectos de la lingüística cuantitativa y la modelización matemática de la información UNLP. (2008) Disponible en: <https://memoria.fahce.unlp.edu.ar/tesis/te.350/te.350.pdf>. Fecha de consulta: 07/02/22
- [7] Robredo, J. (2019). Otimização dos processos de indexação dos documentos e de recuperação da informação mediante o uso de instrumentos de controle terminológico. *Ciência Da Informação*, 47(1). Recuperado de <http://revista.ibict.br/ciinf/article/view/4431>. Fecha de consulta: 07/02/22.
- [8] Martínez Méndez, F. (2004). Recuperación de información: modelos, sistemas y evaluación. Disponible en: <http://eprints.rclis.org/16262/1/libro-ri.PDF>. Fecha de consulta: 07/02/22.
- [9] Salton, G., Wong, A., & Yang, C. S. (1975). A Vector Space Model for Information Retrieval. *Communications of the ACM*, 18(11), 613–620. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.446.5101&rep=rep1&type=pdf>. F. consulta: 07/02/22.
- [10] Zazo Rodríguez A. y otros. (2002). Recuperación de información utilizando el modelo vectorial. U. de Salamanca. Disponible en: <http://eprints.rclis.org/13963/1/zazo2002recuperacion.pdf>. Fecha de consulta: 07/02/22.
- [11] Robaldo, L. y otros. Compiling regular expressions to extract legal modifications. 250. 133-141. 10.3233/978-1-61499-167-0-133. (2012).

Aplicación de Machine Learning sobre imágenes utilizadas en proyectivas

Olivar Matías¹, Sattolo Iris¹, Panizzi Marisa¹

¹Escuela Superior de Ingeniería, Informática y Ciencias Agroalimentarias.

Universidad de Morón.

Cabildo 134, Buenos Aires, Argentina.

matiasolivar@outlook.com, iris.sattolo@gmail.com, marisapanizzi@outlook.com

Resumen

Esta línea de investigación se centra en el estudio y aplicación de técnicas de aprendizaje automático, específicamente, redes neuronales convolucionales y aprendizaje profundo para la resolución de problemas sobre reconocimiento de patrones en imágenes y videos. En esta investigación se pretende analizar las imágenes que se obtienen al aplicar técnicas proyectivas en el campo de la psicología, que permitan a los psicólogos acceder al conocimiento de la subjetividad.

Palabras clave: Machine Learning, Deep Learning, psicología, técnicas proyectivas.

Contexto

Este artículo sintetiza uno de los trabajos de investigación realizado en el marco de una tesina de grado de la carrera Licenciatura en Sistemas de la Escuela Superior, de Ingeniería, Informática y Ciencias Agroalimentarias, de la UM. En esta investigación se aplicarán técnicas de aprendizaje profundo en el área de psicología. Específicamente, se trabajará con técnicas proyectivas. Las técnicas proyectivas son métodos que usan los psicólogos para acceder al conocimiento de la subjetividad, el cual es la expresión teórica del aparato psíquico [1].

Introducción

En las últimas décadas, la información se ha convertido en la herramienta más valiosa en cualquier ámbito o industria para poder tomar mejores decisiones. En el campo de la psicología, los psicoanalistas utilizan las técnicas proyectivas. Las mismas buscan una descripción que pueda abarcar la personalidad y que posean estímulos más ambiguos. El analista busca que el paciente transfiera emociones reprimidas o no aceptadas hacia otra persona, dibujos u objetos con el fin de poder liberar esa emoción reprimida que se encuentra en el inconsciente de la persona [2].

La Tabla 1 presenta las técnicas proyectivas más utilizadas en el ámbito de la psicología clínica, laboral y forense.

Para esta investigación se seleccionó la técnica denominada “Persona bajo la lluvia”, esta le permite al analista producir un mínimo de ansiedad en el sujeto analizado y evaluar mediante el dibujo realizado, aspectos tanto de personalidad, como de conducta ante situaciones desfavorables o negativas. Puede aplicarse tanto en niños, con el fin de obtener información sobre situaciones traumáticas vividas sin la necesidad de que el paciente pediátrico lo comunique verbalmente. Y también en adultos con la misma

finalidad, o para realizar una evaluación conductual del sujeto ante situaciones de tensión o negativas.

Para la realización de esta, se requieren materiales sencillos, poco tiempo para su administración, pudiendo aplicarse tanto de forma individual como grupal.

Tabla 1. Técnicas proyectivas y su respectivo propósito.

Técnica	Propósito
Persona bajo la lluvia	Externalizar la percepción del individuo bajo condiciones desfavorables [3].
Rorschach	Identificar el funcionamiento psíquico del entrevistado. [4]
Apercepción Temática	Método con el cual el diagnosticado puede conocer los impulsos, sentimientos y conflictos de una personalidad [4].
Familia kinética	Evaluación de la inserción del examinado en la familia y la importancia que se le asigna. Generalmente es usado con niños [4].
Casa, árbol, persona	Técnica orientada a conocer rasgos puntuales de la personalidad [5].
Figura humana	Búsqueda de los sentidos, conflictos, ansiedades e impulsos de la persona [5].

Se le entrega al examinado una hoja A4 y se le pide que dibuje una persona bajo la lluvia. El tiempo promedio de realización oscila entre 10 y 20 minutos, aunque es importante tener en cuenta que la consigna es libre y por lo tanto cada sujeto se toma el tiempo necesario para realizarlo. El aumento o disminución del tiempo de ejecución es indicador de características de la personalidad del sujeto. Esta prueba permite

conocer cómo se defiende el sujeto en situaciones de tensión ambiental. A diferencia de otros test, este solicita al entrevistado que dé dos tipos de respuesta, una gráfica y la otra verbal. La respuesta gráfica permite estudiar indicadores estructurales o expresivos (dimensión, emplazamiento, trazo, presión, borrado, etc.), e indicadores de contenido (orientación de la persona, postura, uso del paraguas, lluvia, identidad, transparencias, etc.) [6].

En la Figura 1 se presenta un ejemplo del test de la persona bajo la lluvia.

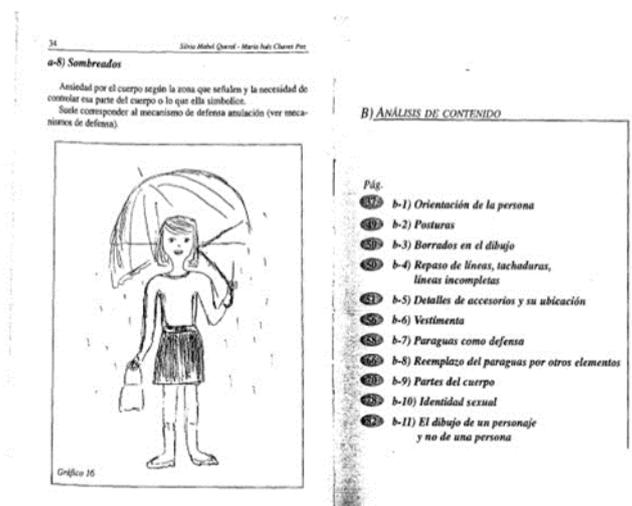


Figura 1. Ejemplo del test persona bajo la lluvia [6].

Estos dibujos podrían ser utilizados para obtener patrones con técnicas de aprendizaje profundo (en inglés, *Deep Learning*).

El aprendizaje automático según Hawkins [7] es la capacidad adquirida por un ordenador de identificar patrones de datos y así de esta forma poder elaborar predicciones.

El aprendizaje automático es una rama de la

inteligencia artificial que estudia sistemas capaces de aprender a realizar una tarea a partir de datos de ejemplo. Es de naturaleza inductiva y comprende técnicas y métodos para realizar clasificación, optimización y predicción, mayormente en dominios en donde los problemas no pueden definirse de forma explícita o no existen soluciones analíticas aplicables. Por estos motivos, las técnicas que presenta resultan adecuadas para el procesamiento de imágenes y otras señales.

Las redes neuronales (en inglés, *Neural Networks* o RN) son modelos de aprendizaje automático que consisten en un grafo de computación no lineal con capacidad de aproximación universal y una gran cantidad de parámetros. Dichos parámetros son ajustados mediante un proceso de optimización llamado entrenamiento a partir de datos de ejemplo. La optimización busca en general minimizar cierta función de error asociada con alguna tarea en particular. Los modelos de aprendizaje automático basados en redes neuronales han permitido mejoras de desempeño muy significativas, estableciéndose en los últimos años como la tecnología base del estado del arte. El progreso se debe a la utilización de un conjunto de técnicas de aprendizaje profundo (Deep Learning). [8]

Deep Learning es utilizado para realizar procesos de Machine Learning empleando redes neuronales artificiales compuestas por varios niveles jerárquicos. En el primer nivel la red aprende patrones simples, y esta información se envía al siguiente nivel de la jerarquía. Este segundo nivel toma la información obtenida en el primero y la combina con nuevos patrones

aprendidos en este, generando información un poco más compleja, la cual es pasada a un tercer nivel, y así sucesivamente. Las técnicas de Machine Learning y Deep Learning proveen gran soporte para el diseño de aplicaciones de visión por computadora o visión artificial [9].

Las redes neuronales convolucionales (en inglés, *Convolutional Neural Network* (CNN)), son un tipo especial de redes neuronales, las cuales se están utilizando para procesar imágenes y reconocer aspectos importantes de ellas para luego realizar predicciones en base a lo aprendido [10]. Según el artículo [11], las CNN fueron elegidas por sobre los demás tipos de redes neuronales profundas, para la realización de estudios relacionados con la medicina.

Dada la evolución de la aplicación de técnicas de Deep Learning se han desarrollado herramientas como Pylearn o Deepmat [12], que permiten agregar valor a este tipo de estudios.

Existen diversas investigaciones del ámbito de la psicología clínica en las cuales se hace uso de Machine Learning o Deep Learning, como en [13], en la cual utilizan un modelo predictivo para el diagnóstico temprano de desórdenes psicológicos en una persona.

En la investigación llevada a cabo por Edward Grant *et al.* [14], se estudia las atribuciones psicológicas de una persona mediante técnicas de CNN, con el fin de lograr predicciones de personalidad y comportamiento. A diferencia de este último estudio, en [15] podemos encontrar una investigación del estado psíquico de una persona que se encuentra bajo tratamiento oncológico. El estudio demuestra que la ansiedad y la depresión son factores muy importantes, tanto para el tratamiento, como para

evitar suicidios. Esta investigación utiliza redes neuronales profundas para lograr un modelo predictivo sobre el estado de la salud mental de dichos pacientes.

De los trabajos analizados no se encontraron estudios específicos que utilicen técnicas de Machine Learning o Deep Learning en el análisis de técnicas proyectiva.

Líneas de Investigación, Desarrollo e Innovación

Esta investigación propone la utilización de técnicas de Machine Learning para el análisis de técnicas proyectivas usadas en psicología. Específicamente, el uso de las imágenes que se utilizan en la técnica de “persona bajo la lluvia”

La pregunta de investigación (PI) que guiará la propuesta es:

El aprendizaje profundo, ¿puede ayudar a clasificar los dibujos producidos por las personas a las cuales se les pide completar el test de la persona bajo la lluvia?

Se orientarán los esfuerzos a responder distintos objetivos que ayudarán con el desarrollo de este trabajo.

- 1) Revisar la literatura existente sobre aplicaciones de Deep Learning orientados a la psicología. Se empleará el método de investigación, Mapeo sistemático de la literatura (en inglés, *systematic mapping study* o SMS) de acuerdo con las directrices de Kitchenham *et al* [16].
- 2) Relevar y obtener un data-set que contenga gráficos que se utilizan como ejemplos en el test propuesto.

- 3) Seleccionar los indicadores de contenido para procesar los datos.
- 4) Proponer el modelo de aprendizaje profundo que se utilizará en el trabajo.
- 5) Validar la propuesta con expertos en el campo de la psicología que apliquen estos tipos de test.

Dado que es un trabajo que se inicia recientemente, se pretenden lograr los resultados que se detallan a continuación:

- a) Resultados académicos, tesis de grado.
- b) Producción Científica: se presentarán los avances de la investigación en CACIC¹ 2022 e InNGENIO² 2022).

Formación de Recursos Humanos

Las asignaturas de Tesis de las carreras de Informática de la UM fomentan la investigación en los estudiantes. Estos mismos proponen un tema de su interés y desarrollan su trabajo de tesis bajo la supervisión de los docentes de la cátedra. En este caso, este trabajo es desarrollado por un estudiante de grado de la carrera Licenciatura en Sistemas de la UM.

Referencias

1. Celener, G. (2006). Las Técnicas Proyectivas. Buenos Aires: Lugar Editorial.
2. Soave, M., (2016). Manual de Técnicas Proyectivas. Buenos Aires: Editorial Brujas
3. Pérez Zambón, S., (2015) EL TEST DE PERSONA BAJO LA LLUVIA. UNA NUEVA PERSPECTIVA DE ANÁLISIS Subjetividad y Procesos Cognitivos, Vol. 19, N° 1, Pág. 200-227, ISSN impreso: 1666-244X, ISSN electrónico: 1852-7310

¹ Congreso Argentino de Ciencias de la Computación (CACIC).

² Congreso Latinoamericano de Ingeniería. Link: <http://fundacioniai.org/ingenio/>

4. Mirotti, A. (2008). Introducción al Estudio y Práctica de las Técnicas Proyectivas. Brujas
5. Soave, M., Huespe, T., Villagra, L., Ferrer, C., Muszio, S., Saavedra, M., Chavez, L., (2016). Manual de técnicas proyectivas. Brujas. Argentina
6. Querol, S., Chaves Paz, M., (2005). Test de Persona Bajo la Lluvia: adaptación y aplicación. Lugar
7. Hawkins, J., (2004). On Intelligence. Times Books.
8. Gerón, A., (2017). "Hands-On Machine Learning with scikit Learn & TensorFlow". O'reilly.
9. Budumax, N. (2017). "Fundamentals of Deep Learning". O'reilly.
10. LeCun, Y., (2015). "Deep Learning".
11. Le Lu, Y., (2017). Deep Learning and Convolutional Neural Networks for Medical Image Computing. Suiza.
12. Kustikova, P., (2016). A survey of deep Learning methods and software tools for image recognition.
13. Aryal, Y., Maag A., and Gunasekera, N., (2020) "Application of Machine Learning algorithms in diagnosis and detection of psychological disorders", 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications, Australia.
14. Grant, E., Sahm, S., Zabihi, M., Gerven, M., (2015) "Predicting and visualizing psychological attributions with a deep neural network", Paises Bajos.
15. Shafiei, S., Lone, Z., Elsayed, A., Identifying mental health status using deep neural network trained by visual metrics. Transl Psychiatry.
16. Kitchenham, B., Budgen, D., y Brereton, P., (2015), Evidence-Based Software Engineering and Systematic Reviews, 1 st. Edition ed., Chapman and Hall/CRC

Tecnologías de la información facilitadoras para la interoperabilidad de software en Gobierno Abierto

Roxana Martínez, Diego Wiernik, Gastón Axel Lacuesta,
Ignacio Ezequiel Rondan, Federico Rivarola, Luciano Dodaro

Instituto de Tecnología (INTEC)
Universidad Argentina de la Empresa (UADE)
Lima 775, Ciudad Autónoma de Buenos Aires, Argentina

ing.roxana.martinez@gmail.com; dwiernik@uade.edu.ar

{glacuesta, irondan, ferivarola, ldodaro}@uade.edu.ar

RESUMEN

El concepto de Gobierno Abierto se encuentra en pleno auge, tanto a nivel nacional como internacional. Este cambio de paradigma en la gestión pública surge por la necesidad de tener un mejor acercamiento con el ciudadano a través del uso de las TIC (Tecnologías de la Información y la Comunicación), y, además, ejercer el derecho al acceso de la información pública.

La gestión de los aspectos relacionados a la interoperabilidad permite conocer como las plataformas digitales pueden intercambiar datos, y a su vez comprender el tratamiento de éstos, de forma eficiente y estandarizada para compartir recursos de información y realizar transacciones entre los mismos. A través de procesos coordinados y automatizados se pueden lograr intercambios simples y seguros de datos, fomentando así, una gran base de conocimiento público y accesible.

Este trabajo permitirá conocer el estado y situación actual de las tecnologías utilizadas en Gobierno Abierto que permiten brindar interoperabilidad entre los organismos gubernamentales y los ciudadanos. Por otra parte, conocer las mejores prácticas en aspectos de niveles técnicos para el tratamiento de los datos públicos que son asociados (relacionados entre sí) y visualizados en los portales estatales nacionales desde enfoques del diseño y arquitectura de software.

Palabras clave: Datos Abiertos, Gobierno Abierto, Interoperabilidad en informática, Diseño y Arquitectura de componentes para intercambio de datos.

CONTEXTO

El presente trabajo es parte del proyecto denominado “Tecnologías de la Información facilitadoras para la Interoperabilidad en Gobierno Abierto”, que tuvo inicio en el mes de octubre 2021. Este proyecto pertenece a la línea de investigación de Ingeniería de Software (IS) del Instituto de Tecnología (INTEC). Las actividades del Instituto están directamente relacionadas con las carreras grado y posgrado de la Facultad de Ingeniería y Ciencias Exactas de la Universidad Argentina de la Empresa (UADE). Alguna de las tareas de este proyecto, se realizan en los laboratorios disponibles en los UADE Labs, edificio tecnológico inaugurado en 2010. El proyecto es financiado y evaluado por la Coordinación de Investigaciones, unidad que depende de la Secretaría Académica de la Universidad, tiene una duración de 2 años, y cuenta con la participación de docentes y estudiantes de grado y posgrado en diversas carreras.

1. INTRODUCCIÓN

Cada vez son más las personas que utilizan las Tecnologías de la Información como el medio de intercambio de datos, a través de plataformas de software y la infraestructura

adecuada entre los ciudadanos y los organismos estatales y privados, lo que permite favorecer a un contexto de participación y colaboración pública entre estos. Para ello, es importante conocer el término de Gobierno Abierto que “es una doctrina política que surge a partir de la adopción de la filosofía del movimiento del software libre a los principios de la democracia. Este paradigma tiene como objetivo que la ciudadanía colabore en la creación y mejora de servicios públicos y en el robustecimiento de la transparencia y la rendición de cuentas” [1]. Por lo que el Estado Nacional debe incorporar las técnicas para gestionar dicha cantidad de datos con un diseño y arquitectura tecnológica adecuada.

Uno de los pilares sobre los que se sustenta el concepto de Gobierno Abierto es el de la interoperabilidad [2], el cual ha evolucionado desde su primera definición que se refería a la habilidad de dos o más sistemas de intercambiar información y el uso de esta [3]. En la actualidad, la interoperabilidad puede definirse como un complejo concepto que incluye diferentes dimensiones: técnicas, semánticas, organizacionales, legales y políticas, y modelos de gobernanza [4] y, además, la inclusión del enfoque de datos abiertos públicos que se pueden reutilizar y redistribuir sin ninguna restricción [5]. Estas dimensiones marcan que, si bien la interoperabilidad es un proceso tecnológico en su esencia, este no es el único aspecto para considerar [2], requiriendo un especial foco en los equipos de trabajo interdisciplinarios dentro de las administraciones para su correcta implementación [6].

Si se realiza un desglose de las dimensiones mencionadas, se pueden encontrar dos grupos. Por un lado, los aspectos legales y políticos que conforman el marco en el cual la interoperabilidad se desarrolla, generando los acuerdos que aseguren el cumplimiento de metas políticas en lo individual y general, y garantizando la normativa vigente [2]. Por otro, se encuentran las dimensiones relacionadas la información en sí misma: la estructura semántica que ésta debe tener para que sea interpretada

correctamente, es decir, los recursos tecnológicos para que los datos sean efectivamente intercambiados y las condiciones organizacionales que deben generarse para que la colaboración se realice en forma efectiva [2] [7].

Profundizando dentro de la dimensión técnica, se puede destacar que para favorecer a la interoperabilidad y, por extensión, a la generación de frameworks sobre el cual se sustentará el gobierno abierto, es necesario que la información cuente con al menos 5 (cinco) características fundamentales [8]: debe ser abierta, por definición; debe tener la calidad y cantidad necesaria; debe poder ser utilizable por todos; debe ser liberada para mejorar la gobernanza; y finalmente, debe ser liberada para mejorar la innovación. Esto no solamente evidencia la importancia de esta dimensión [4], sino también las dificultades y desafíos a enfrentar. Con la aplicación de estos conceptos, se busca mejorar aspectos de calidad en los datos abiertos públicos [11] e interoperabilidad de éstos, con el fin de replicar las mejores prácticas de los países líderes en este contexto y contar con un modelo de calidad estándar como la ISO 25012 [10], para generar gran valor y aprovechamiento por parte de los diferentes usuarios [9].

“La interoperabilidad y la utilización de estándares abiertos permiten la compatibilidad entre distintas tecnologías y ahorrar en costos de desarrollo o contratación de servicios. Además, facilita la colaboración entre organismos al mismo tiempo que fomenta la transparencia en la Administración Pública y la reducción de la dependencia de oferentes” [12]. Alguno de los recursos a tener presentes en este contexto son posibles propuestas de soluciones, que utilizan estándares abiertos para maximizar la compatibilidad con otras plataformas gubernamentales, para aumentar la transparencia y facilitar la colaboración, algunos de estos recursos son: estándares de la W3C [13], estándares abiertos para APIs estilo REST [14], Pautas Técnicas de Interoperabilidad de Sistemas (Anexo II RES 19 E-2018) [15], Estándares de Servicios

Digitales y Herramientas de la Plataforma Digital del Sector Público Nacional [16] [17], Herramientas para la construcción colaborativa de contenidos de la Plataforma Digital del Sector Público Nacional, como por ejemplo la plataforma mundial de intercambio de datos, MagicBox [18], la plataforma INSIDE [19], que facilita la gestión de archivos, registros, y expedientes de manera interoperable, o bien la propuesta del algoritmo para conectar automáticamente portales de datos abiertos con un nodo central llamado Pydatajson [20], entre numerosas herramientas más [21].

Como se explicó anteriormente, diversos organismos estatales ofrecen una gran cantidad de fuentes de datos de varios temas gubernamentales con criterios preestablecidos en sus portales y brindan datasets que son utilizados como insumo fundamental de información y servicios, es por lo que, los datos son una parte integral de los esfuerzos para ofrecer oportunidades y poder realizar un gobierno más transparente, participativo y eficiente. A su vez, es necesario que el diseño y construcción de los servicios digitales se base en soluciones [22] que logren una adecuada comunicación e intercambien información, es decir, que interoperen.

2. LÍNEAS DE INVESTIGACIÓN DESARROLLO

Este proyecto pertenece a la línea de investigación de Ingeniería de Software (IS) del Instituto de Tecnología (INTEC). Los ejes principales del tema que se están investigando en referencia a las actividades de I+D+i (Investigación, Desarrollo e Innovación) son:

- Analizar las falencias en cuestiones tecnológicas actuales de interoperabilidad.
- Elaborar el estado situación de la interoperabilidad en contexto de datos abiertos y públicos.
- Desarrollar una guía de las mejores prácticas en aspectos de niveles técnicos para el tratamiento de los datos públicos que son asociados (relacionados entre sí) y visualizados en los portales estatales nacionales.

- Diseñar y desarrollar una propuesta de prototipo de software utilizando las mejores técnicas de diseño y arquitectura de software orientadas es este contexto.

Fundamentación del proyecto:

Desde el enfoque en Audiencia Científica:

Los resultados (parciales y finales) de este proyecto serán transferidos para contribuir con nuevo conocimiento a la comunidad académica y, a la sociedad en general en el ámbito nacional e internacional. Estas contribuciones se encontrarán plasmadas en forma de publicaciones.

Desde el enfoque en Audiencia como

Institución: La audiencia interna la constituyen los estudiantes y docentes de la carrera de Ingeniería Informática y afines, pudiendo también incorporarse integrantes de diferentes carreras interesados en las cuestiones del gobierno abierto en esta temática.

Desde el enfoque Social: El proyecto corresponde a un abordaje desde la disciplina del Software a aspectos del ámbito de la gestión pública, tratándose de cuestiones de creciente preocupación. Su base fundamental surge por la necesidad de tener un mejor acercamiento tanto con el ciudadano como con las organizaciones a través del uso de las TIC, y, además, que estos puedan ejercer el derecho al acceso de la información pública.

3. RESULTADOS OBTENIDOS/ESPERADOS

Enfoque del proyecto:

Se buscará brindar nuevos enfoques especialmente en el diseño y construcción de tecnologías para la interoperabilidad, entendida como la capacidad de comunicar y transferir datos entre sistemas de información para posibilitar el intercambio de información y conocimiento entre ellos. Por las características del proyecto guarda fuerte vínculo con conceptos de Ingeniería de Software, Diseño de Software y Arquitectura de Software.

Objetivos principales:

Identificar los principales aspectos tecnológicos que son facilitadores para llevar a cabo la interoperabilidad en el contexto de Gobierno Abierto y brindar buenas prácticas a niveles técnicos en lo que concierne a esta temática. Esto conduce a comprender las nociones que son fundamentales en este nuevo paradigma gubernamental desde un aspecto de diseño y arquitectura de software.

Objetivos específicos esperados:

Relevar los conceptos fundamentales en esta temática que se relacionan con el marco de interoperabilidad; Analizar y comparar los antecedentes a nivel internacional sobre dicho tema; Elaborar el estado situación de la interoperabilidad en contexto de datos abiertos y públicos; Comprender e identificar el marco regulatorio legal de la interoperabilidad en la Administración Pública; Relevar las distintas tecnologías a nivel técnico, que son utilizadas en aspectos de interoperabilidad; Estudiar las técnicas más relevantes en contextos de interoperabilidad en la Administración Pública (diseño y arquitectura de software); Definir las ventajas y desventajas de la implementación de la interoperabilidad; Relevar los tipos de datos utilizados (sensibles, abiertos, confidenciales, entre otros) en este entorno; Identificar las mejores prácticas en aspectos de niveles técnicos para el tratamiento de los datos públicos que son asociados (relacionados entre sí) y visualizados en los portales estatales nacionales.

Metodología y Técnicas:

En los que respecta a las técnicas de investigación, se realizarán enfoques cualitativos que permitan obtener estudios enfocados a:

- Observación de contextos de Gobierno Abierto.
- Investigación bibliográfica: Con el fin de apoyar y sustentar el trabajo investigativo. Relevar los estudios previos y contrastarlos a la actualidad.
- Estudios de Casos en la República Argentina y países extranjeros: Para

recolectar información pertinente al estudio y obtener una representación lo más similar a la realidad relevada.

- Narrativa y métodos de visualización.
- Estudio de Técnicas de interoperabilidad y conexión interna de datos abiertos en organismos públicos nacionales desde el diseño y arquitectura de software.
- Análisis y encuadre del framework de la metadata en datos abiertos públicos gubernamentales.

4. FORMACIÓN DE RECURSOS HUMANOS

Este proyecto se compone por 2 (dos) docentes de grado y posgrado que cuentan con varios años de experiencia en el dictado de clases en varias Universidades de Argentina y más de 18 años de experiencia en el ambiente laboral en tecnología informática, también tienen estudios de posgrado: uno de ellos Magíster en Tecnología Informática de la Universidad Abierta Interamericana (UAI), y por otro lado, se encuentra a la espera de respuesta por parte del jurado (actualmente en proceso de revisión) de su tesis doctoral en Ciencias Informáticas de la Universidad Nacional de La Plata (UNLP), y otro docente Magíster en Gestión de Servicios de Tecnología y Telecomunicaciones de la Universidad de San Andrés, y además, Especialista en Redes y Servicios en Telecomunicaciones de la Universidad de Buenos Aires (UBA).

El equipo también cuenta con la participación de estudiantes de grado y de posgrado de carreras de la rama de Informática y Sistemas de la UADE (Universidad Argentina de la Empresa).

En relación directa con la línea de I+D+i presentada para el proyecto, los miembros del equipo se encuentran en realización de: 3 trabajos finales de carreras afines en la UADE. Otro de los integrantes del equipo es estudiante de la carrera de Licenciatura en Gestión de Tecnología de la Información en UADE, quién aporta su especial conocimiento desde su experiencia laboral, ya que actualmente se desempeña en el área

de datos públicos del Honorable Senado de la Nación de la República Argentina.

5. BIBLIOGRAFÍA

[1] Sosteniblepedia.org. “*Gobierno Abierto*”. Disponible en:

https://www.sosteniblepedia.org/index.php?title=Gobierno_abierto

[2] D'Agostino, S. (2011). Desarrollo de un Framework para la Interoperabilidad en Gobierno Electrónico (Doctoral dissertation, Universidad Nacional de La Plata).

[3] IEEE Standard Glossary of Software Engineering Terminology. (1990). IEEE Std 610.12. *IEEE Computer Science*, 42.

[4] Jiménez, C. E., Solanas, A., & Falcone, F. (2014). E-government interoperability: Linking Open and Smart Government. IEEE Computer Society, 22-24.

[5] Oviedo, E., Mazón, J. N., & Zubcoff, J. J. (2013). Hacia un modelo de calidad de datos para portales de datos abiertos. In XXXIX Latin American Computing Conference (CLEI), Naiguata (pp. 1-8).

[6] Sánchez, C. (2019). Interoperabilidad en la Gestión Pública.

[7] Colpaert, P., Compennolle, M. V., Vocht, L. D., Dimou, A., Sande, M. V., Verborgh, R., . . . Mannens, E. (2014). Quantifying the Interoperability of Open Government Datasets. IEEE Computer Society, 50 - 56.

[8] UK Government Cabinet Office. (18 de June de 2013). G8 Open Data Charter and Technical Annex. Disponible en:

www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex

[9] de Colombia, G. (2020). Calidad e Interoperabilidad de los datos abiertos del Gobierno de Colombia. Disponible en: https://herramientas.datos.gov.co/sites/default/files/2020-11/A_guia_de_estandares_final_0.pdf

[10] ISO 25012 (2008). “*Ingeniería de software - Requisitos de calidad y evaluación de productos de software (SQuaRE) - Modelo de calidad de datos*”. Disponible en: <https://www.iso.org/obp/ui/es/#iso:std:iso-iec:25012:ed-1:v1:en>

[11] Martínez, R. et al. (2021). Metrics proposal to measure the quality of governmental datasets. IEEE Latin America Transactions, Vol. 100. ISSN 1548-0992.

[12] Argentina.gov.ar (2022). Decálogo Tecnológico ONTI. Utiliza estándares abiertos y soluciones interoperables. Disponible en:

<https://www.argentina.gov.ar/jefatura/innovacion-publica/ssetic/ont/decologo-tecnologico/utiliza-estandares-abiertos-y-soluciones-interoperables>

[13] W3C. “*Standards*”. Disponible en:

<https://www.w3.org/standards/>

[14] OpenAPI (2022). “*OpenAPI Initiative*”. Disponible en: <https://www.openapis.org/>

[15] Resolución, Secretaría de Modernización Administrativa (2018). “*Pautas Técnicas de Interoperabilidad de Sistemas*”. Disponible en:

https://www.argentina.gov.ar/sites/default/files/ont/ont/res_19_2018_anexo_ii_if_2018_09344892_apn_ssga_mm.pdf

[16] Argentina.gov.ar (2022). “*Contenidos Digitales*”. Disponible en:

<https://www.argentina.gov.ar/contenidosdigitales>

[17] HM Government (2012). “*Open Standards Principles. For software interoperability, data and document formats in government IT specifications*”. Disponible en: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/459074/Open-Standards-Principles-2012.pdf

[18] BID (2021). “*MagicBox*”. Disponible en:

<https://code.iadb.org/es/herramientas/magicbox>

[19] BID (2021). “*INSIDE*”. Disponible en:

<https://code.iadb.org/es/herramientas/inside>

[20] BID (2021). “*Pydatajson*”. Disponible en: <https://code.iadb.org/es/herramientas/pydatajson>

[21] BID (2021). “*Reutiliza estas herramientas*”. Disponible en:

<https://code.iadb.org/es/funcionalidad/intecambio-e-interoperabilidad-de-datos>

[22] Araujo, S., & Vargas, M. P. (2020). La interoperabilidad en el marco del Gobierno Digital. Conocimiento Libre y Licenciamiento (CLIC), (22).

MODELO PREDICTIVO PARA EVUALAR EL RIESGO POTENCIAL DE EXISTENCIA DE EPIDEMIA

Aristides Dasso, Ana Funes

Departamento de Informática / Facultad de Ciencias Físico-Matemáticas y Naturales /
Universidad Nacional de San Luis

Ejército de los Andes 950, D5700HHW San Luis, Argentina

+54 (0) 266 4520300, ext. 2126

{ arisdas, afunes}@unsl.edu.ar

RESUMEN

Como parte integral de una línea de investigación en el desarrollo de modelos de evaluación de sistemas complejos, y considerando que la determinación de la existencia de una posible epidemia, o inclusive de una pandemia, en un área o población determinada, es de máxima importancia, y que una alerta temprana es una herramienta de prevención para que las autoridades competentes puedan comenzar a tomar las medidas preventivas necesarias, es que presentamos los objetivos, lineamientos generales y resultados esperados de un trabajo de investigación sobre la creación de modelos de evaluación de epidemias.

Esta investigación tiene como objetivo la creación, puesta a punto y aplicación de modelos que permitan obtener indicadores del nivel alcanzado en la evaluación de alertas tempranas ante la posibilidad de presencia de una epidemia.

La metodología a seguir para el desarrollo de dichos modelos de evaluación está basada en la aplicación del método Logic Score of Preference (LSP) [15]. Asimismo, tomamos como referencia para la creación del modelo, algunas de las recomendaciones de organismos tales como el Center for Disease Control (CDC) de los EEUU [9], el European Centre for Disease Prevention and Control (ECDPC)

[11], así como de otras publicaciones como [1], [7], [8], [9], [16].

Palabras clave: Epidemia. Pandemia. Métodos de Evaluación Multicriterio. Evaluación cuantitativa. Logic Score of Preference (LSP).

CONTEXTO

El trabajo de investigación aquí presentado se encuentra enmarcado dentro del ámbito de la Universidad Nacional de San Luis, ejecutándose dentro de una de las líneas de investigación del Proyecto de Ciencia y Técnica PROICO 03-2020 “Ingeniería de Software: Estrategias de Desarrollo, Mantenimiento y Migración de Sistemas en la Nube”, dirigido por el Dr. Daniel Riesco. El mismo se encuentra acreditado con evaluación externa y financiamiento de la Universidad Nacional de San Luis.

INTRODUCCIÓN

La determinación temprana de la posible presencia de una pandemia en un territorio es una tarea de suma importancia para los organismos de salud y gubernamentales.

Se pueden encontrar, en la literatura, múltiples propuestas para evaluar dicha posibilidad, tales como las de Carter C. Price and Adrienne M. Propp [8], J. Li et al.[16], Pluchino et al. [1], entre otras.

Muchos de estos trabajos están basados en anteriores epidemias de influenza como, por ejemplo, la herramienta IRAT [10] o el trabajo de C. Reed et al. [7], quienes han creado un marco de trabajo (framework) para evaluar los efectos en la salud pública de una epidemia emergente. Por su parte, el Centro Europeo para la Prevención y Control de Enfermedades (European Centre for Disease Prevention and Control) ha examinado dicho framework y ha concluido que se trata de un paso adelante en el análisis realizado con un acercamiento único [11].

Para evaluar los efectos de una enfermedad en un área dada y la probable transformación en una epidemia, deben considerarse un número de factores, atributos o parámetros tales como la severidad de la enfermedad, el número de personas afectadas, entre otros. A su vez, estos factores pueden verse afectados por otros parámetros tales como la población, la estación del año, el índice R_0 [18], la tasa de fatalidad, la inmunidad subyacente de la población, etc. Todo ello puede servir para detectar el preludio de una epidemia o pandemia o simplemente de un pequeño brote.

Que un organismo de salud pública posea un modelo que permita evaluar la posibilidad cierta de que una enfermedad pueda ser una amenaza grave a la salud pública es claramente de la máxima importancia para los servicios de salud y, si bien contar con esos modelos constituye una necesidad importante, la construcción de los mismos no es una tarea sencilla.

Múltiples aspectos deben ser considerados en esta tarea, teniendo en cuenta diversos criterios de decisión, como los antes mencionados. Sin embargo, la adopción de un método sistemático de desarrollo de modelos de evaluación multicriterio como el método Logic Score of Preferences (LSP) puede ayudar a lidiar con esta complejidad y poner orden en todo el proceso. Debido justamente a la complejidad de muchos de los frameworks propuestos en la literatura, obtener un indicador global podría ser una tarea dificultosa, aún cuando muchos de esos frameworks adoptan un método de evaluación

simple como lo son los métodos aditivos. Sin embargo, gracias al método iterativo y jerárquico de agregación de características propuesto por el LSP, los atributos identificados pueden ser adecuadamente tratados para producir, no sólo un indicador global y una conclusión con respecto a una posible epidemia, sino también obtener información parcial o focalizada.

Asimismo, cabe destacar que LSP presenta una serie de ventajas frente a otras técnicas aditivas de evaluación multicriterio:

- Permite construir una Estructura de Agregación (EA), a partir de un conjunto de operadores de una lógica continua (CLP), que refleja las características del razonamiento humano [14], permitiendo modelar condiciones de: simultaneidad, neutralidad y reemplazabilidad. De tal forma, LSP puede comportarse como una técnica compensatoria o no compensatoria, permitiendo al decisor elegir los criterios que pueden ser compensados y cuales presentan restricciones al respecto.
- Provee un conjunto de operadores de agregación, que van desde un grado máximo de reemplazabilidad, pasando por la neutralidad hasta llegar a la máxima simultaneidad [13]. En consecuencia, brinda la base para evaluar una preferencia con distintos grados de intensidad.
- En el contexto de la toma de decisiones, no todas las evaluaciones se definen por el grado de conjunción o por el grado de disyunción entre factores. Algunas veces, la preferencia entre factores, no puede establecerse como completamente reemplazable o simultánea. En este sentido, LSP proporciona las bases para construir operadores más complejos que responden a esta necesidad, caso del operador compuesto CPA (Conjunctive Partial Absorption).

Asimismo, para la selección de los factores o parámetros intervinientes en el modelo de decisión se analizarán diversas propuestas, dadas como bases para frameworks, que se encuentran en la literatura, como, por ejemplo,

los propuestos en Manoj Gambhir et al. [17] que sirven para predecir epidemias.

LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

La línea de investigación en la que se enmarca el trabajo presentado, es parte de una investigación sobre la construcción de modelos de evaluación de sistemas complejos, que viene desarrollándose desde hace tiempo en el marco de un Proyecto de Ciencia y Técnica de la Universidad Nacional de San Luis, donde se han obtenido resultados plasmados en diversas publicaciones (ver, por ejemplo, [2][3][4]).

Los principales ejes del presente trabajo de investigación giran en torno a los siguientes puntos:

- Conceptualización del problema de decisión en cuestión, analizando y definiendo los factores específicos que deben ser tenidos en cuenta en este contexto particular.
- Modelización, a partir de los factores intervinientes para la toma de decisión, aplicando un método de evaluación multicriterio, que permita la obtención de un valor numérico útil como indicador para la toma de decisión en un contexto de epidemia.
- Validación del modelo propuesto por contrastación con datos obtenidos de epidemias previas.

RESULTADOS Y OBJETIVOS

El objetivo principal de este trabajo es elaborar y validar un modelo que permita, a través de múltiples factores, decidir sobre la posibilidad de la existencia de una epidemia.

Objetivos Específicos

- Definición de una jerarquía de factores: El primer paso en el desarrollo consiste en realizar la descomposición del problema en una jerarquía con tantos niveles como sean necesarios. Esto brinda una visión del problema y sirve

para identificar un conjunto de atributos o factores intervinientes en la toma de decisión. A partir de la revisión de la literatura, se ha observado que existen diversas propuestas que servirían de base para la creación de dicha jerarquía (Árbol de Preferencias en LSP).

- Normalización los atributos: Una vez que se han identificado los atributos, es necesario normalizarlos. Esto, en LSP, es la etapa donde se definen funciones llamadas criterios elementales que transforman los valores de los atributos a valores en el intervalo [0,100].
- Clasificación de los atributos: Una vez que se tiene la jerarquía de atributos, es necesario clasificarlos en obligatorios, opcionales y deseables para poder, luego, crear la EA.
- Elaboración de la función multicriterio o EA: El objetivo de esta etapa es agregar preferencias elementales y parciales, por medio de los operadores LSP, para obtener una preferencia global que sirva para decidir, ante una determinada situación, si los valores asignados a los atributos están o no indicando la presencia de una epidemia. Esto es posible debido al método de evaluación adoptado, que va más allá de un simple método de evaluación aditivo, permitiendo que puedan darse no solo mayor o menor peso a distintos ítems, sino admitiendo la posibilidad de modelar condiciones de simultaneidad, neutralidad y reemplazabilidad.
- Validación del modelo: A partir de datos obtenidos de epidemias previas, se realizarán evaluaciones que sirvan para contrastar los resultados obtenidos con el modelo contra los de la realidad.

FORMACIÓN DE RECURSOS HUMANOS

La línea de investigación sobre evaluación de sistemas complejos que se viene desarrollando en el marco del Proyecto de Ciencia y Técnica PROICO 03-2020 "Ingeniería de Software: Estrategias de

Desarrollo, Mantenimiento y Migración de Sistemas en la Nube”, ha dado lugar a numerosas publicaciones y tesis de grado y de posgrado. En este sentido, creemos que la propuesta de investigación acá presentada seguirá dando sus frutos, tanto en publicaciones nacionales e internacionales (p.e. [5], [6]) como en la formación de recursos humanos (dos tesis de maestría presentadas más una tesis de maestría en ejecución). Asimismo, de momento, se ha encarado la posibilidad de la ejecución de una nueva tesis de maestría basada en los objetivos que aquí nos hemos propuesto.

REFERENCIAS Y BIBLIOGRAFÍA

- [1] A. Pluchino, A. E. Biondo, N. Giuffrida, G. Inturri, V. Latora, R. Le Moli, A. Rapisarda, G. Russo & C. Zappalá. “A novel methodology for epidemic risk assessment of COVID-19 outbreak”. *Scientific Reports* (2021) 11:5304 <https://doi.org/10.1038/s41598-021-82310-4>. www.nature.com/scientificreports
- [2] Ana Funes, Aristides Dasso, Germán Montejano, Daniel Riesco. “A SAMM-based model for assessing Cybersecurity Implementations”, actas de CoNaIISI 2018, 29 y 30 de Noviembre de 2018, Mar del Plata, Buenos Aires, Argentina.
- [3] Aristides Dasso y Ana Funes, “Threat and Risk Assessment Using Continuous Logic”, *Encyclopedia of Organizational Knowledge, Administration, and Technologies*, 1st. edition. IGI Global. Aceptado para su publicación en 2020.
- [4] Aristides Dasso, Ana Funes, Germán Montejano, D. Riesco, R. Uzal, Roberto, N. Debnath; “Model Based Evaluation of Cybersecurity Implementations”. ITNG 2016. Las Vegas, Nevada, USA, 11-13 abril 2016. In S. Latifi (ed.), *Information Technology New Generations, Advances in Intelligent Systems and Computing* 448. DOI: 10.1007/978-3-319-32467-8_28. Springer International Publishing, Switzerland 2016.
- [5] C. Gallardo, A. Funes, H. Ahumada. “Soporte para la Medición y Evaluación de la Accesibilidad al Contenido en Aplicaciones Web”, *Anales de ASSE 2019 (JAIIO 2019)*, Salta, Argentina. pp. 56-70.
- [6] C. Gallardo, A. Funes. “Un Modelo para la Evaluación de la Calidad de la Accesibilidad al Contenido Web”, *CONAISI 2015*, Bs. As., Argentina.
- [7] C. Reed and M. Biggerstaff and L. Finelli and L. Koonin and D. Beauvais and A. Uzicanin and A. Plummer and J. Bresee and S. Redd and D. Jernigan. “Novel Framework for Assessing Epidemiologic Effects of Influenza Epidemics and Pandemics”. *Emerging Infectious Diseases*, 2013, volume 19, pages 85 – 91.
- [8] Carter C. Price, Adrienne M. Propp. “A Framework for Assessing Models of the COVID-19 Pandemic to Inform Policymaking in Virginia”. Published by the RAND Corporation, Santa Monica, Calif. © Copyright 2020 RAND Corporation.
- [9] CDC “Pandemic Severity Assessment Framework (PSAF)”. November 3, 2016. <https://www.cdc.gov/flu/pandemic-resources/national-strategy/severity-assessment-framework.html>
- [10] CDC Summary of Influenza Risk Assessment Tool (IRAT) Results, Pandemic Influenza (Flu), https://www.cdc.gov/flu/pandemic-resources/monitoring/irat-virus-summaries.htm?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fflu%2Fpandemic-resources%2Ftools%2Firat-virus-summaries.htm.7/8/2021
- [11] European Centre for Disease Prevention and Control “Epidemiologic framework for assessing the severity of influenza epidemics and pandemics”. 11 Jan 2013
- [12] Farzaneh Sadat Tabataba, Prithwish Chakraborty, Naren Ramakrishnan, Srinivasan Venkatramanan2, Jiangzhuo Chen2, Bryan Lewis2 and Madhav Marathe1,2 A framework for evaluating

- epidemic forecasts. *BMC Infectious Diseases* (2017) 17:345. DOI 10.1186/s12879-017-2365-1
- [13] J. Dujmović, & H. Nagashima. 2006. LSP Method and its Use for Evaluation of Java IDEs. *International Journal of Approximate Reasoning*. 41, 3-22.
- [14] Jozo Dujmović, Andness and Orness as a Mean of Overall Importance. *Fuzzy Systems (Fuzz-IEEE)*, 2012 IEEE International Conference on , 10-15 June 2012. 2012. 1-6.
- [15] Jozo Dujmović. “Soft Computing Evaluation Logic. The LSP Decision Method and Its Applications”. © 2018 John Wiley & Sons, Inc.
- [16] JunHua Li, Pradeep Ray, Holly Seale, Raina MacIntyre AN E-HEALTH READINESS ASSESSMENT FRAMEWORK FOR PUBLIC HEALTH SERVICES – PANDEMIC PERSPECTIVE. 2012 45th Hawaii International Conference on System Sciences. Jan. 4 2012 to Jan. 7 2012. Maui, Hawaii USA. ISBN: 978-0-7695-4525-7.
- [17] Manoj Gambhir, Catherine Bozio, Justin J. O’Hagan, Amra Uzicanin, Lucinda E. Johnson, Matthew Biggerstaff, and David L. Swerdlow. “Infectious Disease Modeling Methods as Tools for Informing Response to Novel Influenza Viruses of Unknown Pandemic Potential”. *Clinical Infectious Diseases®* 2015;60(S1):S11–9. Published by Oxford University Press on behalf of the Infectious Diseases Society of America. DOI: 10.1093/cid/civ083
- [18] Wikipedia, “Ritmo reproductivo básico”. https://es.wikipedia.org/wiki/Ritmo_reproductivo_básico

Desarrollo de middleware y aplicación cliente para sistema de miniboyas ambientales

Zaradnik, Ignacio; Dominguez, Facundo; Kumvich, Augusto; Lupi, O.Daniel;
Caccaviello, Diego.

Laboratorio de Inteligencia Ambiental Departamento de Ingeniería e Investigación Tecnológica,
Universidad Nacional de La Matanza. Buenos Aires, Argentina

izaradnik@unlam.edu.ar; dominguez@unlam.edu.ar; kumvich@gmail.com; olupi@unlam.edu.ar;
dcaccaviello@unlam.edu.ar

RESUMEN

El presente trabajo detalla el desarrollo y la implementación de un middleware y la aplicación cliente de un sistema de miniboyas ambientales. Se comienza planteando la importancia del agua como recurso natural y los medios para asegurar su calidad. A continuación, se describe el hardware utilizado en la miniboya y los antecedentes asociados al presente trabajo. Finalmente, se realiza una breve explicación de la arquitectura del software implementado, se detallan sus principales elementos y las consideraciones tenidas en cuenta en el desarrollo del middleware y de la aplicación cliente.

Palabras Clave: Middleware, Internet de las cosas, Ecosistema Acuático, SQL, MQTT.

CONTEXTO

En el marco del Laboratorio de Inteligencia Ambiental del Departamento de Ingeniería e Investigación Tecnológica de la Universidad Nacional de La Matanza, se está trabajando desde hace algunos años en aplicaciones de Internet de las Cosas (IoT) [1][2][3]. El presente trabajo es parte de lo realizado en el marco del proyecto “Internet de las Cosas en Miniboyas Ambientales”, el cual se ha desarrollado entre comienzos del 2020 y fines del 2021. Este trabajo se financió con fondos provenientes del Programa de Incentivos para Docentes Investigadores de la Secretaría de Políticas Universitarias (PROINCE).

1. INTRODUCCION

En la actualidad existe una creciente preocupación por el deterioro del medioambiente y por el impacto que

determinadas actividades humanas pueden causar sobre él. En especial lo que afecta a los recursos naturales, destacándose el agua, que es un elemento básico para la vida. Según la UNESCO (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura), el agua dulce es el recurso más importante para la humanidad, ya que abarca todas las actividades sociales, económicas y ambientales. Es una condición para toda la vida en nuestro planeta, un factor habilitador o limitante para cualquier desarrollo social y tecnológico, una posible fuente de bienestar o miseria, cooperación o conflicto [4]. En nuestro país, datos de AySA (Agua y Saneamientos Argentinos) indican que el 13% de la población no tiene acceso a agua potable [5]. El saneamiento inadecuado del agua para consumo humano es una de las causas de múltiples enfermedades y muertes a nivel mundial [6]. Por esta razón, determinar la calidad de las fuentes de agua que se utilizan para el consumo humano y de animales resulta indispensable, pudiendo así evitar las enfermedades y muertes resultantes del consumo de agua con el incorrecto saneamiento. Para asegurar la calidad del agua se consideran parámetros físicos, químicos y biológicos. Estos parámetros se fijan de manera diferenciada según los usos a los que se va a destinar el recurso (consumo humano, riego, industria, ganadería, vida acuática). En función de esta problemática es que se planteó el desarrollo de un sistema de miniboyas para el monitoreo de los ecosistemas acuáticos.

2. LINEAS DE INVESTIGACION y DESARROLLO

El objetivo general de este trabajo es desarrollar, implementar y estudiar los resultados del uso del sistema experimental de miniboyas ambientales. Para ello se investigaron: los distintos parámetros a medir para determinar la calidad del agua y los sensores asociados, la electrónica necesaria para acondicionar las señales de los sensores y para su procesamiento, las tecnologías de comunicaciones para la transmisión de los datos recolectados y los distintos medios para implementar un middleware y la aplicación cliente [7][8][9]. El presente trabajo se enfoca en este último punto.

3. RESULTADOS OBTENIDOS

3.1. Hardware del sistema

En la figura N°1 se puede observar el diagrama de bloques de la miniboya, mientras que en la figura N°2 se puede ver la maqueta construida con fines de evaluar los aspectos mecánicos de la misma. El módulo GNSS (Sistema Global de Navegación por Satélite) empleado es el SL869V2, mientras que el módulo de conectividad celular es el UL865-NAD, ambos de la empresa Telit. El microcontrolador utilizado es el ATSAM4S16 de la empresa Microchip, el cual se encuentra integrado en una placa de desarrollo Xplained. Como alimentación se empleó un pack de baterías de Níquel-Metalhidruro (Ni-MH) de 4,8 V y 2100 mAh. En lo que respecta a los sensores, se consideraron los siguientes: oxígeno disuelto (SEN0237-A); conductividad (DFR0300-H) y pH/temperatura (SEN0249).

3.2. Antecedentes

En paralelo al desarrollo de los programas presentados en este trabajo se analizaron distintas alternativas de computación en la nube: IBM Cloud Solutions, Microsoft Azure Cloud, Google Cloud, Telit device wise y Digi Remote Manager. Como consecuencia de este análisis, se desarrolló una interfaz gráfica basada en la opción Telit device wise [9]. La elección de esta opción se fundamentó en que: ofrece el uso de la plataforma sin costo y sin limitaciones en el desarrollo de la interfaz (solo

en la cantidad de dispositivos conectados y el tiempo de permanencia de los datos); no limita el tiempo de uso de la plataforma; permite la utilización de mapas para la geolocalización; al trabajar con un módulo celular de Telit (como se realizó), la integración de éste a la plataforma es mucho más sencilla y el proveedor brinda soporte para la implementación. Con este análisis y el desarrollo implementado, se logró adquirir un mayor entendimiento de las distintas alternativas para futuros proyectos.

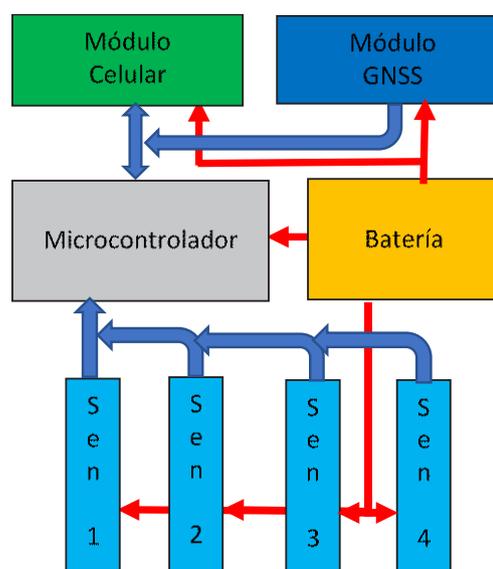


Figura N°1. Diagrama en bloques del sistema.

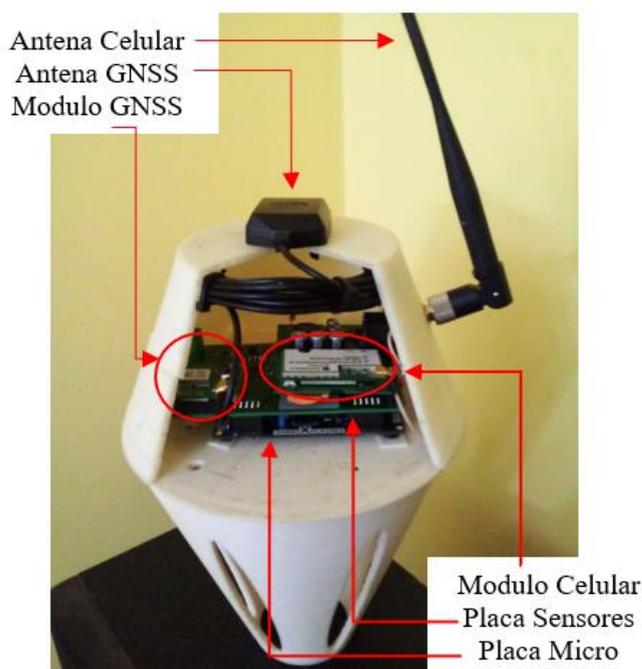


Figura N°2. Sistema experimental.

3.3. Desarrollo

3.3.1. Descripción General

En la figura N°3 se puede ver la arquitectura del software implementado (middleware y cliente). Dicho software se implementó en servidores propios, es decir, no se usó ningún servicio de computación en la nube. El middleware implementado consta de tres elementos: un broker MQTT, un intérprete de datos y una base de datos MySQL.

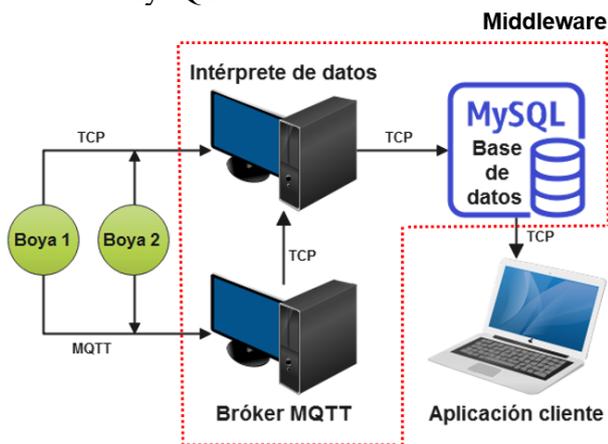


Figura N°3. Arquitectura del Software.

Periódicamente, los valores de los sensores y la posición brindada por el módulo GNSS son leídos por el microcontrolador, quien empaqueta dichos datos junto con un identificador único de la boya y un CRC (código de redundancia cíclica) para la detección de errores. Luego, los datos empaquetados son transmitidos al middleware a través de la conexión celular, por medio del protocolo TCP o MQTT. En el primer caso la comunicación se establece directamente con el intérprete de datos, mientras que en el segundo la miniboya publica sus datos con un tópico específico (INTERPRETE) en el broker MQTT. Al encontrarse el intérprete de datos suscripto a dicho tópico en el mismo broker (ya que se encuentra funcionando como un cliente MQTT), este va a recibir los datos transmitidos por las miniboyas. A la trama de datos recibida, ya sea a través del protocolo TCP o MQTT, se le verifica la integridad a través del CRC integrado en la misma. Si esta fuese correcta, se confirma la recepción satisfactoria a la miniboya ya sea a través de un mensaje por medio del protocolo TCP o publicando el mismo en el broker MQTT en un tópico particular

(ESTACIONES), al cual las miniboyas se encuentran suscriptas. Validado el mensaje recibido, se extraen sus datos de este y se almacenan en la base de datos, quedando la información disponible para ser accedida por la aplicación cliente.

3.3.2. Broker MQTT

Como broker MQTT se utilizó Eclipse Mosquitto™, un servidor de mensajes de código abierto (con licencia EPL/EDL) que implementa las versiones 5.0, 3.1.1 y 3.1 del protocolo MQTT. Su instalador se puede descargar de <https://mosquitto.org/download/>. Luego de la instalación, fue necesario configurar el broker para permitir la conexión de clientes que se encuentren en otro dispositivo. Para ello, se debió editar el archivo “mosquitto.conf”, ubicado en el directorio de instalación (por defecto “C:\Program Files (x86)\Mosquitto”), agregando las siguientes líneas para indicar el puerto de conexión, y habilitar la conexión de clientes anónimos, sin que tengan que autenticarse:

- listener 1883
- allow_anonymous true

Por último, fue necesario permitir la conexión del archivo “mosquitto.exe” en el firewall de Microsoft Windows, creando la regla de entrada y salida correspondiente.

3.3.3. Intérprete

El intérprete se realizó con la versión de prueba LabWindows/CVI 2015 SP1 cuyo instalador, se encuentra disponible de manera gratuita en la página web de National Instruments, [10]. Para la comunicación del intérprete con la miniboya a través del protocolo TCP, se emplearon librerías estándar provistas en la instalación del LabWindows/CVI. Mientras que, para la comunicación con la base de datos, se usó la librería libmysql versión 5.7.31.0, provista por MySQL, disponible para su descarga de <https://dev.mysql.com/downloads/mysql/5.7.html>. Finalmente, para la comunicación con el broker MQTT se debió integrar la librería libmosquitto al LabWindows/CVI, disponible con la instalación del broker junto a su

documentación

(<https://mosquitto.org/man/libmosquitto-3.html>). La figura N°4 presenta la interfaz gráfica de la aplicación interprete. A través de esta, se podrá configurar la conexión a la base de datos (dirección IP de servidor de base de datos, usuario y contraseña), el puerto TCP al cual se podrán conectar las miniboyas y los parámetros para la conexión al broker MQTT (dirección IP del broker, puerto de comunicación, tópico al cual subscribirse y tópico en el cual publicar).

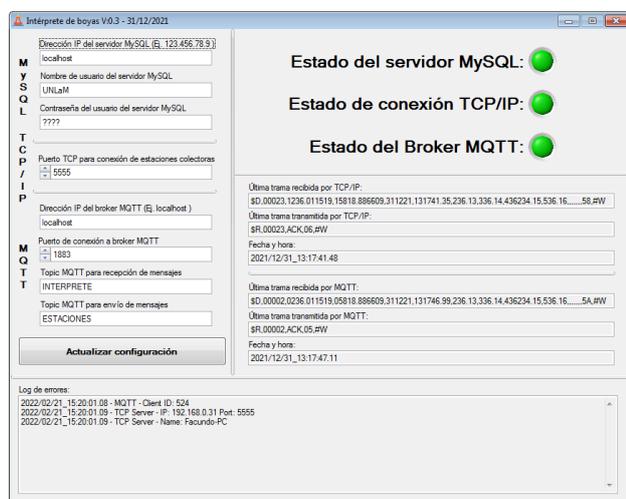


Figura N°4. Interfaz gráfica del interprete.

3.3.4. Base de datos

Como base de datos se empleó MySQL (<https://www.mysql.com/>), el cual es un sistema de bases de datos relacional, multihilo y multiusuario con licencia GNU GPL. Para la presente aplicación solo se ha implementado una tabla con los siguientes parámetros:

- **IDLog:** Identificación única del registro
- **IDDispositivo:** Identificación del dispositivo que realizó la medición
- **Latitud:** Latitud del punto en donde se realizó la medición
- **Longitud:** Longitud del punto en donde se realizó la medición
- **Fecha:** Fecha de la medición
- **Hora:** Hora de la medición
- **S1:** Valor del sensor 1 (Oxígeno disuelto)
- **S2:** Valor del sensor 2 (pH)
- **S3:** Valor del sensor 3 (Conductividad)
- **S4:** Valor del sensor 4 (Temperatura)

- **Fecha Carga:** Fecha en la cual se cargó el registro en la base de datos
- **Hora Carga:** Hora en la cual se cargó el registro en la base de datos.

3.3.4. Aplicación cliente

Al igual que el intérprete, la aplicación cliente fue realizada con LabWindows/CVI. La figura N°5 presenta la pantalla principal de la aplicación.

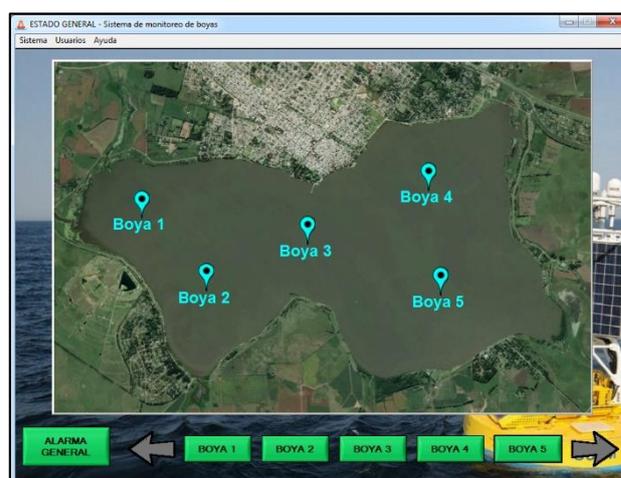


Figura N°5. Pantalla principal de la aplicación cliente.

La pantalla principal consta de un mapa con la ubicación de cada una de las boyas (actualmente es una imagen fija georreferenciada, pero está previsto la inclusión de un motor de mapas). Debajo de esta, se encuentran botones que permiten el acceso a la información de cada una de las boyas. En la figura N°6 se presenta la pantalla de cada una de las boyas, en donde se puede ver la fecha, la hora y los valores de la última medición recibida, así como también un gráfico histórico con los valores registrados.

3.4. Conclusiones

Se logró la implementación de un middleware y la aplicación cliente de un sistema de miniboyas. Si bien el sistema desarrollado es de carácter experimental, a través del mismo se han logrado habilidades en el desarrollo de hardware, firmware (asociado a la miniboya), uso de servicios en la nube, desarrollo de software con la integración de servicios distribuidos (broker MQTT y base de datos

MySQL) y el desarrollo de aplicaciones móviles, las cuales no fueron tratadas por cuestiones de espacio.

A futuro se prevé continuar la investigación de la temática de control de calidad de ecosistemas acuáticos, integrando nuevas variables a monitorear (micro/nano plásticos y cianobacterias), implementando mejoras en la interfaz gráfica (como la integración de motores de mapas), incluyendo seguridad middleware y usando librerías de análisis de datos para obtener información de los datos obtenidos.



Figura N°6. Pantalla de información de boya.

4. FORMACION DE RECURSOS HUMANOS

El ámbito de este proyecto permitió tanto la formación grupal del equipo de trabajo, así como la individual de cada uno de sus miembros. La formación grupal buscó generar conocimiento en la temática de aplicaciones de Internet de las cosas y monitoreo de la calidad de los ecosistemas acuáticos. En relación las formaciones individuales, se enumeran a continuación: Ignacio Zaradnik la gestión de grupos de trabajos, Diego Caccaviello la revisión bibliográfica y la elaboración de estados del arte, Facundo Dominguez el diseño de software, Diego Turconi el diseño de aplicaciones de sistemas embebidos y Augusto Kumbich el desarrollo del hardware.

5. REFERENCIAS

[1] Canziani; Gomez; Lupi; Nassipián; Slawiski; Turconi; Zaradnik, “Plataforma de conexión de Redes Eléctricas

Inteligentes a Internet de las Cosas” en el Congreso Argentino de Sistemas Embebidos 2014. ISBN 978-987-45523-27.

[2] Bernis; Turconi; Benacerraf; Dominguez; Lupi; Zaradnik; Rzepa, “Sistema de seguimiento de dosimetría personal”. VII congreso de microelectrónica aplicada 2016. ISBN: 978-987-733-068-7.

[3] Lupi; Zaradnik; Turconi; Dominguez, “Sistema de visualización de precios para supermercados”. Congreso Argentino de Sistemas Embebidos 2017 (Case 2017), Buenos Aires, Argentina. ISBN: 978-987-46297-3-9.

[4] UNESCO (n.d.). “Water Security”. Extraída el 06/03/2022 desde <https://en.unesco.org/themes/water-security>

[5] Pablo Bereciartua (2019), “Los desafíos del agua en la Argentina: el desarrollo del Plan Nacional del Agua”. Extraída el 14/05/2021 desde https://www.argentina.gob.ar/sites/default/files/02_pb_los_desafios_del_agua_en_la_argentina.pdf.

[6] Médicos sin fronteras (n.d), “Agua y Saneamiento. ¿Por qué se debe proporcionar agua y saneamiento?”. Extraída el 14/05/2021 desde <https://www.msf.org.ar/conocenos/actividades-medicas/agua-y-saneamiento>

[7] O.Lupi, I.Zaradnik, M.Canziani. “Estado del arte de los sistemas de monitoreo de calidad de agua”. Revista Digital del Departamento de Ingeniería (Reddi). Vol.5 número 2 (2020). Publicado 31/12/2020.

[8] O.Lupi, D.Turconi, J.Slawiski. “Monitoreo de Ecosistemas Acuáticos”. Revista Digital del Departamento de Ingeniería (Reddi). Vol.7 número 2 (2021). Publicación pendiente.

[9] O.Lupi, I.Zaradnik, A.Agüero, C.Behar, L.Lanzilliotti, M.Vázquez. “Interfaz gráfica en la nube para monitoreo de miniboyas ambientales”. 5to Congreso Argentino de Ingeniería. Libro de Resúmenes (10/2021).

[10] “LabWindows” (n.d.). Extraída el 06/03/2022 desde <https://www.ni.com/es-cr/support/downloads/software-products/download.labwindows-cvi.html#306964>

ANÁLISIS E IMPLEMENTACIÓN DE CIUDADES INTELIGENTES

Rocío A. Rodríguez, Pablo M. Vera, Claudia G. Alderete, Mariano G. Dogliotti

Centro de Altos Estudios en Tecnología Informática (CAETI)
Facultad de Tecnología Informática
Universidad Abierta Interamericana (UAI)
Montes de Oca 745, Ciudad Autónoma de Buenos Aires, Argentina

{rocioandrea.rodriguez, pabломartin.vera, claudia.alderete, mariano.dogliotti}@uai.edu.ar

RESUMEN

En la actualidad la tecnología es parte de la vida cotidiana de las personas tanto para trabajo, comunicaciones, esparcimiento; las empresas invierten y desarrollan servicios tecnológicos que les facilitan la vida a las personas mientras consumen sus productos y servicios. Hoy en día es común tener aplicaciones en el smartphone para pedir un vehículo para traslado, solicitar comida a domicilio y un sinfín de otras cosas. Los gobiernos no pueden escapar a esta tendencia y deben actualizarse, hacer uso de la tecnología existente, diseñar recursos y servicios para facilitar las gestiones de sus ciudadanos a la vez que brinda mayor transparencia a sus acciones. Es por eso que surge el concepto de ciudades inteligentes o Smart Cities donde la tecnología se hace presente. Este proyecto de investigación se basa en el estudio de las Smart Cities (sus bases, fundamentos, grado de implementación) con el objetivo de realizar propuestas para facilitar y mejorar la implementación de las mismas.

Palabras clave: Ciudades Inteligentes, TIC, Participación Ciudadana, Gobernanza, Transparencia

CONTEXTO

Esta línea de investigación y desarrollo (I+D) forma parte de los proyectos radicados en el Centro de Altos Estudios en Tecnología Informática (CAETI) de la Universidad Abierta Interamericana (UAI). En este proyecto participan docentes y alumnos tanto

de sede Centro como de la Castelar (ambas en la provincia de Buenos Aires). El proyecto cuenta con financiamiento asignado y una duración de 2 años.

1. INTRODUCCIÓN

Las TIC (Tecnologías de la Información y Comunicación) han cambiado radicalmente la forma de vida de las personas. Muchas áreas fueron mejoradas o enriquecidas con la llegada de las TIC. Es interesante ver como demográficamente se fueron produciendo cambios los cuales implicaron que en la región de América Latina más del 80% de la población está viviendo en áreas urbanizadas [1], “en 1950, apenas el 42% de la población de la región vivía en ciudades” [2]. “La urbanización es una característica de la civilización contemporánea y el proceso del desarrollo urbano en sí está condicionado por varios factores, entre los cuales se encuentran la globalización y el progreso tecnológico” [3]. Esta sobrepoblación requiere pensar en ciudades que puedan brindar adaptarse y ser sostenibles. “En términos generales el concepto de ciudades o territorios inteligentes y sostenibles hace referencia a un uso extensivo y eficiente de las tecnologías disponibles –en particular las TIC– dirigidas a mejorar la calidad de vida de la población [...]”[4].

Existe diversas definiciones para una ciudad inteligente, entre ellas:

- “Una Smart City es un lugar donde la

tecnología cobra vida” [5].

- Una ciudad inteligente es un área geográfica o territorio que se caracteriza por el uso intensivo de las tecnologías con el objetivo, de manera general, de mejorar la calidad de vida de los ciudadanos y el desarrollo sostenible de las ciudades bajo los supuestos de la colaboración y la innovación [6].
- “Una Ciudad Inteligente es aquella que coloca a las personas en el centro del desarrollo, incorpora Tecnologías de la Información y Comunicación en la gestión urbana y usa estos elementos como herramientas para estimular la formación de un gobierno eficiente que incluya procesos de planificación colaborativa y participación ciudadana. Al promover un desarrollo integrado y sostenible, las Smart Cities se tornan más innovadoras, competitivas, atractivas y resilientes, mejorando así las vidas” [2].

En la figura 2 se plantean algunas características básicas que deben poseer las ciudades inteligentes que constituyen un resumen de principios extraídos del BID (Banco Internacional de Desarrollo) para migrar de una gestión tradicional a la gestión de una Ciudad Inteligente [2].

Los distintos gobiernos han planteado estrategias que permitan realizar gradualmente esta migración. Las estrategias varían dependiendo de las características de cada una de las ciudades. Entender las debilidades y fortalezas de cada ciudad como punto de partida es indispensable para entender como se puede evolucionar para lograr esta migración.

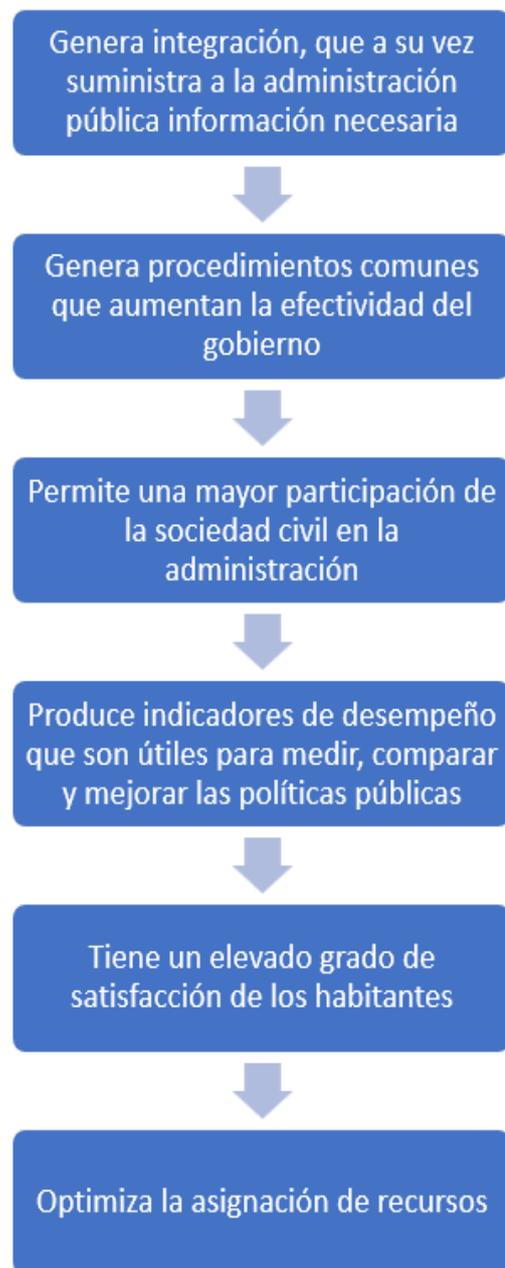


Figura 2. Resumen de las características de una Ciudad Inteligente

En el área de investigación de las ciudades inteligentes existen muchas aristas de trabajo desde la implementación de tecnologías para mejorar aspectos cotidianos ([7], [8], [9]), así como incrementar la participación ciudadana [10], ofrecer a los ciudadanos información de la gestión y transparencia ([11], [12], [13], [14]) ó incluso, proponer un mecanismo para medir su grado de funcionamiento generando ranking de posicionamiento de ciudades [15], [16], [17], [18] [19].

Es importante evidenciar que es un área temática con muchas posibilidades de desarrollo en sus diversas aristas.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Los ejes principales del trabajo son:

- Analizar las implementaciones existentes para Smart Cities y clasificarlas
- Comparar los planes de acción de los gobiernos nacionales de América Latina con respecto a las Smart Cities
- Analizar formas de evaluar el grado en que las ciudades han implementado características de las Smart Cities
- Desarrollar estrategias e implementar recursos que permitan mejorar distintos aspectos propios de las Smart Cities

3. RESULTADOS OBTENIDOS/ESPERADOS

Dado que este proyecto ha comenzado en el presente año aún no hay resultados obtenidos. Pero los resultados a obtenerse serán producto de analizar las soluciones existentes y sus aportes en cada uno de los principios esperados para una ciudad inteligente. Por otra parte, se espera obtener una comparativa sobre ciudades de Latinoamérica. Estos serán los primeros resultados a obtenerse a medida que avance el tiempo de ejecución del proyecto.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo está formado por 4 docentes, 2 de ellos doctores en Ciencias Informática graduados en la Universidad Nacional de La Plata (UNLP) y 1 realizando una maestría

(UAI). Esto implica que 3 de los 4 docentes que componen el grupo tienen estudios de posgrados finalizados o en progreso.

Este proyecto también cuenta con la participación de alumnos de grado y posgrado de la UAI (actualmente en el proyecto se encuentran vinculados 4 alumnos).

En esta área encuentran en realización 1 tesis de doctorado (UNLP) y una tesina de grado (UAI), siendo directores de dichas tesis miembros del equipo de investigación.

5. BIBLIOGRAFIA

- [1] Naciones Unidas, Departamento de Asuntos Económicos y Sociales, División de Población (2014). World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352)
- [2] Bouskela, M., Casseb, M., Bassi, S., De Luca, C., & Facchina, M. (2016). La ruta hacia las smart cities: Migrando de una gestión tradicional a la ciudad inteligente (Vol. 454). Inter-American Development Bank.
- [3] Sikora-Fernández, D. (2017). Factores de desarrollo de las ciudades inteligentes. *Revista Universitaria de Geografía*, 26(1), 135-152.
- [4] Alvarado López, R. A. (2018). Ciudad inteligente y sostenible: hacia un modelo de innovación inclusiva. PAAKAT: revista de tecnología y sociedad, 7(13).
- [5] Góngora, G. P. M. (2015). Revisión de literatura sobre ciudades inteligentes: una perspectiva centrada en las TIC. *Ingeniare*, (19), 137-149.
- [6] TMForumInform (2015) <http://inform.tmforum.org/features-and-analysis/featured/2015/10/peter-sany-smart-cities-are-where-technology-comes-alive/>

- [7] Sánchez, J. A. S. (2017). Redes vehiculares aplicadas a la movilidad inteligente y sostenibilidad ambiental en entornos de ciudades inteligentes (Doctoral dissertation, Universidad de Oviedo).
- [8] Romero Amaya, L. J., & Echeverry Osorio, J. F. (2019). Análisis de producto y modelo de negocio de la Smart City Station.
- [9] Gutiérrez Gutiérrez, J. L. (2021). Algoritmo genético aplicado al recojo de residuos sólidos en el contexto de una smart city. *Revista Cubana de Ciencias Informáticas*, 15(4), 1-14.
- [10] Madrigal-Moreno, S., & Gil-Lafuente, J. (2017). Análisis de los retos del desarrollo sostenible de Barcelona como smart city mediante el estudio de su reputación online. *Cuadernos del CIMBAGE*, 2(19), 29-50.
- [11] Patiño, J. A. (2014). Datos abiertos y ciudades inteligentes en América Latina: Estudio de casos.
- [12] Le Breton, M. A., Girardeau, M., & Bailleul, H. (2021). From Open Data to Smart City Governing Innovation in the Rennes Metropolitan Area (France). *International Journal of E-Planning Research (IJEPR)*, 10(4), 17-38.
- [13] Rose, G., Raghuram, P., Watson, S., & Wigley, E. (2021). Platform urbanism, smartphone applications and valuing data in a smart city. *Transactions of the Institute of British Geographers*, 46(1), 59-72.
- [14] Li, X., Liu, H., Wang, W., Zheng, Y., Lv, H., & Lv, Z. (2022). Big data analysis of the internet of things in the digital twins of smart city based on deep learning. *Future Generation Computer Systems*, 128, 167-177.
- [15] Moreno Alonso, C. (2016). Desarrollo de un modelo de evaluación de ciudades basado en el concepto de Ciudad Inteligente (Smart City) (Doctoral dissertation, Caminos).
- [16] Alvarado-López, R. A. (2020). Ciudades inteligentes y sostenibles: una medición a cinco ciudades de México. *Estudios sociales. Revista de alimentación contemporánea y desarrollo regional*, 30(55).
- [17] Ceballos Zuluaga, A. (2020). Modelo de medición de un sistema de salud en una ciudad inteligente (Doctoral dissertation, Escuela Colombiana de Ingeniería Julio Garavito).
- [18] Alderete, M. V. (2021). Propuesta de un índice de ciudad inteligente para municipios de Argentina. *PAAKAT: revista de tecnología y sociedad*, 11(21).
- [19] Segura-Bonilla, O., & Hernández Milian, J. (2021). Aspectos conceptuales y metodológicos para la construcción de un Índice de Ciudades Inteligentes y Sostenibles.

DESARROLLO DE UNA APLICACIÓN DE GEOFENCING EN ANDROID PARA MONITOREO DE CORREDORES ESTUDIANTILES

Pablo M. Vera, Rocío A. Rodríguez, Hernán A. Viavattene, Cesar D. Delgado,
Esteban A. Carnuccio

Centro de Altos Estudios en Tecnología Informática (CAETI)
Facultad de Tecnología Informática
Universidad Abierta Interamericana (UAI)
Montes de Oca 745, Ciudad Autónoma de Buenos Aires, Argentina

{pabломartin.vera, rocioandrea.rodriguez } @uai.edu.ar
{hernanalberto.viavattene, cesar.delgado, estebanandres.carnuccio }
@alumnos.uai.edu.ar

RESUMEN

Las amplias capacidades de los dispositivos móviles actuales hacen posible el desarrollo de aplicaciones que sirvan para traer mayor seguridad y tranquilidad a sus usuarios. Este proyecto plantea el desarrollo de una aplicación de monitoreo de estudiantes que utilizan ciertos corredores seguros para ir y volver desde su casa hasta el lugar de destino que puede ser el colegio, un club, la casa de un amigo, etc. Esta aplicación es especialmente útil para niños jóvenes que comienzan a movilizarse por su cuenta donde sus padres muchas veces se preocupan por su seguridad. Esta aplicación proporciona un monitoreo no invasivo alertando si se produce un desvío de / los caminos preestablecidos, así como también al entrar o salir a un área determinada en un horario no esperado utilizando técnicas de geofencing. La aplicación consta de un backend de configuración web y una aplicación móvil con la cual se realiza el monitoreo utilizando la geolocalización de los dispositivos móviles.

Palabras clave: Android, Aplicaciones Móviles, APPs, Geofencing, Geolocalización

CONTEXTO

Esta línea de investigación y desarrollo (I+D) forma parte de los proyectos radicados en el

Centro de Altos Estudios en Tecnología Informática (CAETI) de la Universidad Abierta Interamericana (UAI). El proyecto cuenta con financiamiento asignado.

1. INTRODUCCIÓN

Tan sólo en la Ciudad Autónoma de Buenos Aires están definidos 309 senderos escolares [1], estos senderos se encuentran cubiertos por agentes de prevención y cámaras de seguridad que son visualizadas por el centro de monitoreo urbano. Por estos motivos, se consideran a estos senderos escolares como áreas seguras.

Considerando la alta inserción de dispositivos móviles (siendo el sistema operativo Android mayoritario), se ha planificado realizar una aplicación que permita a aquellos padres o tutores que no pueden acompañar a sus niños hasta el colegio, tengan una solución de respaldo que les permita saber si el niño sale de una zona segura, en este caso sale de un corredor escolar. “Los smartphones cuentan con una gran cantidad de sensores que permiten enriquecer las nuevas aplicaciones. Tomando en cuenta su uso masivo y las posibilidades de hardware, componentes y sensores, es importante planificar el desarrollo de aplicaciones innovadoras que aprovechen dicho hardware” [2].

Al definir en una aplicación esos senderos seguros como áreas, se está haciendo uso del concepto de geovallas ó geofencing. Geofencing

“permite el monitoreo remoto de áreas geográficas rodeadas por una cerca virtual (geocerca) y detecciones automáticas cuando los objetos móviles rastreados ingresan o salen de estas áreas” [3].

En este año del proyecto de I+D se está avanzando con las últimas fases del desarrollo de una aplicación Android que permita conocer si un niño ingresa o sale de un área definida como segura y poder lanzar notificaciones a sus padres o tutores. El desarrollo consiste en una aplicación móvil (instalada en el dispositivo del estudiante) y una aplicación web que permite configurar parámetros y alertas. En la figura 1 se grafica el funcionamiento, en la parte superior el recorrido del estudiante que será monitoreado desde el dispositivo móvil, mientras que en la parte inferior se encuentran un conjunto de personas cercanas al estudiante quienes recibirán mensajes de alerta si el estudiante sale de la zona definida como segura, existiendo diversas configuraciones que podrán realizarse desde una aplicación web.

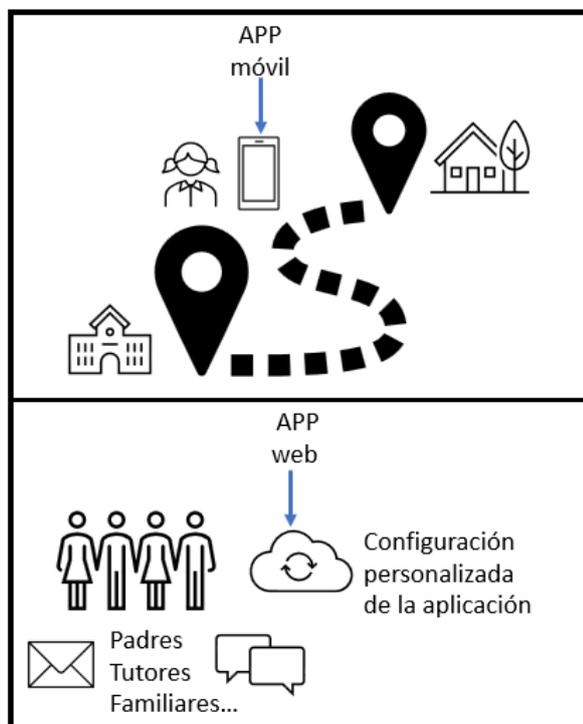


Figura 1. Aplicación Móvil en el dispositivo del estudiante y aplicación web para configuración de personalizaciones y parámetros por parte de un tutor o encargado.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

Este proyecto de investigación busca:

- Hacer uso de la geolocalización y de las técnicas de geofencing para aplicarlos a aplicaciones móviles.
- Diseñar una aplicación de monitoreo no invasiva que consuma pocos recursos.
- Investigar las distintas formas de envío de alertas disponibles para dar mayor flexibilidad al sistema (SMS, WhatsApp, etc)
- Complementar las funcionalidades de las aplicaciones mediante el aprovechamiento de otros sensores disponibles en los dispositivos móviles.

3. RESULTADOS OBTENIDOS/ESPERADOS

El proyecto se encuentra en el proceso final de desarrollo e integración y se espera que, al terminar esta etapa, pueda ser puesto en productivo.

Para realizar su desarrollo se utilizó el Framework de Android Studio, a través del lenguaje de programación Java. Como segundo instrumento, se generó una aplicación web desarrollada con el entorno de desarrollo Visual Studio 2019, utilizando con lenguaje de desarrollo una combinación entre MVC Core y Razor. Para la base de datos, se seleccionó Microsoft SQL 2017 como sistema de gestión, y Transact-SQL como lenguaje de desarrollo. Esta aplicación web podrá ser utilizada a través de cualquier browser y que servirá como asistencia al usuario para poder configurar la aplicación móvil. Dentro de la aplicación web, se podrán realizar las siguientes acciones:

- Generación de áreas de detección de Geofencing: Para la generación de esta funcionalidad, se decidió que era conveniente emplear las APIs nativas de Google de Geofencing [2]. En la figura 2, se muestra a modo de ejemplo una captura de pantalla con un área definida.

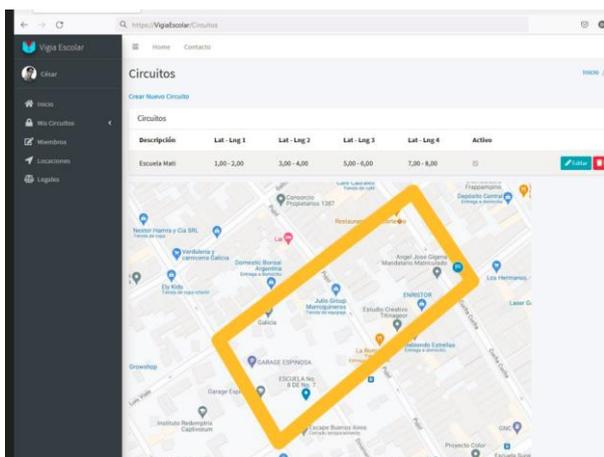


Figura 2. Pantalla de generación de áreas

La funcionalidad de creación de áreas permite no solo determinar las coordenadas permitidas por dichas áreas, sino que también permite configurar dirección origen y destino, tipo de alerta (en el caso de que la persona salga del área), horario en donde el área está habilitada y quien es la persona que debería estar dentro de esa área. Dentro de los tipos de alertas permitidas por la aplicación web, estas podrían ser:

- Envío de email al usuario administrador;
- Envío de mensaje de texto al usuario de la aplicación web ó aplicación móvil.
- Envío de ultima ubicación del usuario móvil al usuario web.
- Generación y administración de usuarios/miembros: Será la aplicación web la cual

permitirá la creación de usuarios, así como también las acciones que pueden realizar los mismos (ver figura 3). Los distintos perfiles que existen son los siguientes:

- Administrador: Será el encargado de crear áreas, usuarios, locaciones. Tendrá acceso a todo el sistema. También será el que reciba las notificaciones en caso de configurarlas.
- Usuario móvil: Podrá ver los caminos configurados para el en la aplicación móvil, y podrá también interactuar con el administrador también a través de la aplicación móvil.
- Usuario web: Podrá ingresar a todas las funcionalidades de la aplicación, pero solo tendrá permisos de lectura.

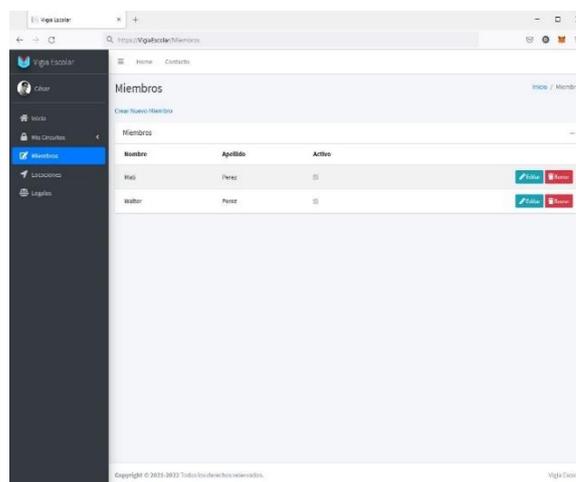


Figura 3. Pantalla de generación y administración de usuarios/miembros

- Generación y administración de locaciones: Una de las partes fundamentales para la generación de áreas de geofencing son las locaciones. Las locaciones serán los lugares

permitidos y rastreables por el usuario administrador, y serán utilizados como punto de referencia para generar las áreas y verificar si la persona se encuentra en esa ubicación en el horario especificado (ver figura 4).

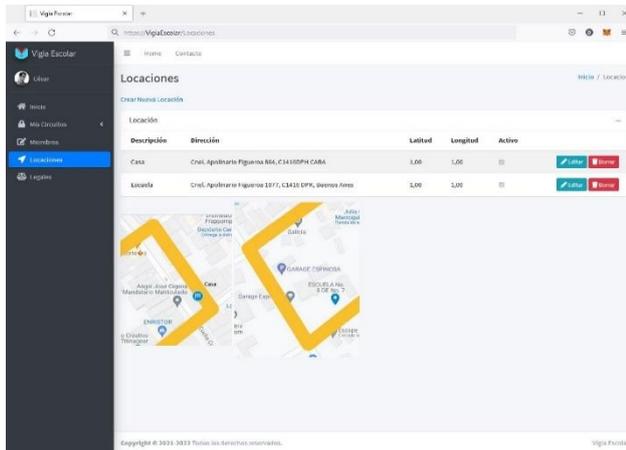


Figura 4. Generación y administración de locaciones

- Visualizar áreas de Geofencing: Como se explicó anteriormente, tanto la aplicación Android como la aplicación móvil utilizarán la API de google para manejar la funcionalidad de Geofencing. En el caso de la aplicación móvil, la misma buscará la información cargada en la aplicación web y mostrará las áreas creadas por pantalla utilizando dicha API. Cada una de las áreas pueden ser de distintos tamaños. Además de ser identificable por un código alfanumérico. Lo cual permite determinar, en qué área se encuentra el usuario y enviar un mensaje de alerta cuando eso suceda. Dichas notificaciones ocurren, cuando el individuo ingrese o salga de estos sectores. En la figura 5 se puede visualizar la aplicación, donde se muestra un mapa con múltiples áreas de Geofencing de diferentes tamaños.

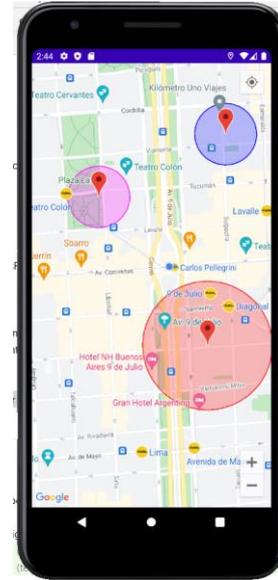


Figura 5. Aplicación con Múltiples Áreas de Geofencing

- Detección de alejamiento de ruta establecida: Una de las utilidades desarrolladas en la aplicación, consiste en establecer un camino que deba seguir el individuo y alertar en caso de que este se desvíe del mismo. Para poder definir la ruta a seguir y poder graficarla en la aplicación, se empleó las APIs de OSRM Directions [5]. Este servicio es Open Source, y es gratuito. Se analizó usar también las APIs de Google Directions [6], pero por cuestiones de limitaciones en su uso, se determinó conveniente emplear OSRM.

Durante el desarrollo del monitoreo de ruta fue necesario analizar cómo se determinaba cuando el usuario debía seguir o no determinado camino. Para ello se decidió, usar dos áreas de Geofencing de activación de ruta. Estas consistían en un área de inicio y otra de fin. De forma tal, que cuando el usuario ingresa dentro del área inicial de una ruta, se activa el monitoreo de seguimiento de dicho

camino. Por ende, si el usuario se aleja varios metros de este, se produce un mensaje de alerta. Informando dicha situación. Posteriormente, el seguimiento de esa ruta se desactiva cuando el usuario ingresa al área de fin, que se encuentra al final del camino. Esta herramienta, resulta de utilidad para hacer uso de un corredor seguro. En otras palabras, cuando un usuario tiene que ir desde un punto A al punto B, siguiendo solo un camino determinado. De manera que, si se desvía, se genera una notificación de alerta. Esta funcionalidad se puede observar en la figura 6.

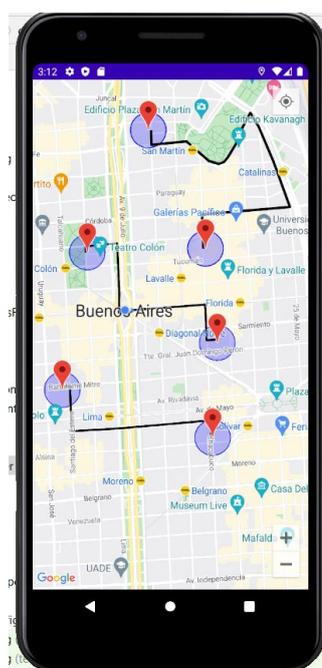


Figura 6. Múltiples Corredores Seguros

Como se puede observar en la figura 6, en la aplicación se pueden generar múltiples recorridos seguros. En donde por defecto, cada uno de ellos se encuentran desactivados. De tal forma, de no estar siendo controlados. Por lo que, un camino es activado, únicamente cuando el usuario ingresa al área de inicio. Cuando esto sucede, se va a monitorear únicamente esa

ruta. Ignorando las demás. Luego esta se desactivará cuando el usuario llegue a su área de finalización.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo está formado por 5 personas, docentes de grado, postgrado y alumnos.

En el área de dispositivos móviles se encuentran en realización, 2 tesis de maestría y 1 tesina de grado en la UAI (Universidad Abierta Interamericana).

5. BIBLIOGRAFÍA

- [1] Buenos Aires Ciudad, “Buscador de Senderos Escolares”. Disponible en: <https://www.buenosaires.gob.ar/justiciayseguridad/senderos-escolares>
- [2] P. Vera, R. Rodríguez, and C. Delgado, “Geofences Application Development for assisting people through monitoring”, LAJC, vol. 9, no. 1, pp. 98-107, Jan. 2022.
- [3] Reclus, F., & Drouard, K. (2009, October). Geofencing for fleet & freight management. In 2009 9th International Conference on Intelligent Transport Systems Telecommunications, (ITST) (pp. 353-356). IEEE.
- [4] Android Developers, “Crea y supervisa geovallas”. Disponible en: <https://developer.android.com/training/location/geofencing>
- [5] OSRM, 2021. “OSRM API Documentation” Disponible en: <http://project-osrm.org/docs/v5.22.0/api/#general-options>
- [6] Android Developers. “The Directions API overview” Disponible en: <https://developers.google.com/maps/documentation/directions/overview>

Question Answering aplicado a la Web Semántica. Predicción de la respuesta esperada.

Matias Oyarzun and Sandra Roger

email: `matias.oyarzun@est.fi.uncoma.edu.ar`
`roger@fi.uncoma.edu.ar`

Grupo de Investigación en Lenguajes e Inteligencia Artificial
Departamento de Teoría de la Computación - Facultad de Informática
UNIVERSIDAD NACIONAL DEL COMAHUE

Resumen

El Procesamiento de Lenguaje Natural es uno de los campos más desafiantes que se tiene en la actualidad. Dentro de éste, una de las áreas que surgen naturalmente es aquella que incluye los Sistemas de Búsqueda de Respuestas (*Question Answering - QA*), cuyo objetivo consiste en dar respuestas correctas o concretas automáticamente a preguntas formuladas por el ser humano en lenguaje natural, evitando traer documentos u otros tipos de fuentes de información extensa.

Al momento de realizar este tipo de sistemas, se hacen presente múltiples dificultades. Esto se debe a que el lenguaje natural es ambiguo y por lo tanto puede ser interpretado de diversas formas. Se hace principalmente evidente durante la interpretación de preguntas, pues basta con malinterpretar el tipo de la misma para generar resultados erróneos. Aquí surge la importancia de poder determinar correctamente lo esperado por la pregunta para poder procesarla, lo que se denomina como *Question/Answer Classification*. Gracias a la evolución de la *Web Semántica*, gran parte de la información disponible en la web se encuentran en forma de bases de conocimientos (*Knowledge Bases - KB*) para ser utilizados, lo que permite minimizar la posibilidad de existencia de ambigüedades, facilitando así el trabajo necesario para el desarrollo de aplicaciones que hagan uso de los datos tal y como es el caso.

Así, el objetivo principal de este plan es la investigación y desarrollo de soluciones basadas en tecnologías de sistemas QA que permitan reducir la búsqueda de información para extraer las respuestas, sobre tecnologías de la Web Semántica a través de herramientas de Lenguaje Natural, lo que contribuye al desarrollo de agentes inteligentes inmersos en la Web. Para esto, se busca en una primera etapa poder realizar una clasificación correcta del tipo de pregunta, lo que permitirá optimizar a las siguientes etapas que abarcan el proceso de búsqueda, pues reduce el espacio de búsqueda e incluso filtrar cualquier tipo de respuesta que no sea apropiada. Una vez que este proceso de clasificación de preguntas se encuentre hecho, se procederá a afrontar el problema de localizar, extraer y presentar al usuario aquella información que desea conocer, mediante una respuesta concreta y de la forma más amigable posible.

Palabras Clave: Sistemas de Búsqueda de Respuestas, Question Answering, Predicción de la Respuesta Esperada, Generalización de Texto, Web Semántica, Procesamiento de Lenguaje Natural.

Contexto

Este trabajo está parcialmente financiado por la UNCo, en el marco del nuevo proyecto de in-

vestigación *Tecnologías Semánticas para el desarrollo de Agentes Inteligentes*. Como así también, lo financia parcialmente el Consejo Interuniversitario Nacional (CIN) con una Beca de Estímulo a las Vocaciones Científicas 2021. El proyecto de investigación tiene una duración de cuatro años y ha comenzado en 2022 y se desarrolla en forma colaborativa con docentes-investigadores de la UNS.

1. Introducción

El rápido aumento de la información y la popularidad del uso de la web se debe a que las personas comienzan a almacenar datos y poner los mismos a disposición del público. Ésto, provoca que el desarrollo de aplicaciones que acceden a estas grandes cantidades de información que se modifican en tiempo real, de manera constante y que posiblemente se encuentren en diferentes formatos, presenten problemas al momento de la exploración de éstos y hace que la búsqueda de información sea una tarea compleja y costosa en términos de tiempo. Pues, muchas de estas aplicaciones ocurren bajo restricciones de tiempo críticas y en intensa interacción con el usuario. Esta dificultad ha motivado el desarrollo de nuevas herramientas de investigación adaptadas, como los sistemas de QA. Por ello, la representación conceptual de los dominios para la generación y extracción de información y conocimiento, es central en la toma de decisiones. De esta manera, los sistemas de QA contribuyen a que el usuario sea capaz de formular una pregunta en lenguaje natural y obtenga una respuesta concreta en lugar de un conjunto de documentos considerados relevantes, como es el caso de los motores de búsqueda.

A su vez, la Web Semántica es un ambiente ideal para el desarrollo de este tipo de agentes, pues ésta busca crear una web de conocimiento en la cual la semántica del contenido es explicitada, permitiendo novedosas aplicaciones que combinan datos heterogéneos para, entre otros objetivos, mejorar la experiencia de los usuarios de acuerdo a sus necesidades. Por ello, lo-

grar que este valioso conocimiento semántico sea accesible y utilizables por los usuarios finales es de principal importancia en los sistemas de QA sobre KB.

Bajo este aspecto, el objetivo general que persigue el proyecto de investigación es el de generar conocimiento especializado en el área de agentes inteligentes y en lo referente a la representación y el uso del conocimiento en sistemas computacionales basados en la web, es decir lo que se ha llamado Web Semántica. Para ello, es necesario profundizar en el estudio de técnicas de representación de conocimiento y razonamiento, tecnologías del lenguaje natural, metodologías de modelado conceptual y mecanismos para la interoperabilidad de aplicaciones, tanto a nivel de procesos como de datos. Se pretende aplicar estos conceptos como soporte para comunidades de desarrollo de ontologías, entre otros.

De esta manera, se busca en una primera instancia la investigación en el área del *Question/Answer Classification* de un sistema QA, pues desempeña un rol importante al momento de determinar las expectativas del usuario. Su objetivo es identificar el tipo de pregunta y, basándose en el mismo, extraer la respuesta esperada de los datos [1]. Ésto permitiría a los sistemas QA identificar de manera más precisa una respuesta adecuada a la pregunta.

El desarrollo del plan de trabajo se realizará en el marco del proyecto de investigación “Tecnologías Semánticas para el desarrollo de Agentes Inteligentes”. En dicho proyecto de investigación se desarrolla una línea de investigación que explora sobre temas afines tanto al análisis y desarrollo de técnicas y herramientas útiles tanto el Procesamiento en Lenguaje Natural como la Generación del lenguaje Natural con el objetivo de dar soporte a los agentes inteligentes en estudio y en la definición de metodologías basadas en técnicas del Lenguaje Natural para el modelado de herramientas de búsquedas de respuestas semánticas. Particularmente, se ha escogido experimentar sobre herramientas de QA que den soporte a la búsqueda de información en el ámbito de la Web Semántica.

2. Línea de Investigación y Desarrollo

El proyecto de investigación *Tecnologías Semánticas para el desarrollo de Agentes Inteligentes* tiene como objetivo general, generar conocimiento especializado en el área de agentes inteligentes que accedan, procesen y recuperen información mediados por tecnologías semánticas.

En este sentido, se desarrolla una línea de investigación que explora sobre técnicas de representación de conocimiento y razonamiento, tecnologías del lenguaje natural, metodologías para la interoperabilidad e integración de datos, y generar nuevos principios, metodologías formales y herramientas basadas en la gestión semántica de los datos. Particularmente, se ha escogido experimentar sobre herramientas de QA que den soporte a la búsqueda de información en el ámbito de la Web Semántica.

El proceso de QA consta de una etapa de análisis de la pregunta, recuperación de los datos relevante de fuentes de conocimiento y la extracción de la información concreta y correcta como respuesta.

El análisis de la pregunta es fundamental. En este sentido, continuando con [2] nos concentramos en una primera etapa en la predicción de la respuesta esperada a partir de la pregunta de entrada. En este sentido, se ha realizado un análisis de las diferentes metodologías y estudio de herramientas disponibles.

Dentro de esta subtarea de QA focalizada en la predicción del tipo de respuesta, existen distintas competencias, una de ella es el desafío denominado SMART *SeMantic Answer Type and Relation Prediction Task*, de la cual se han realizado hasta el momento dos instancias de tales competencias: año 2021¹ y 2020² [3].

La predicción de relaciones para la pregunta es una tarea difícil: algunas relaciones están alejadas semánticamente, a veces los tokens que deciden las relaciones están distribuidas a lo largo de la pregunta, algunas relaciones están

Es posible una clasificación de tipo de respuesta granular con ontologías de Web Semántica populares como DBpedia (~760 clases) y Wikidata (~50K clases). En esta competencia se cuenta con dos tareas principales e independientes: 1) predicción del tipo de respuesta y 2) predicción de un conjunto de relaciones usadas para la identificación de la respuesta correcta.

Se está desarrollando un módulo para la clasificación del tipo de respuesta utilizando aprendizaje automático. La competencia dispone de varios corpus que se pueden utilizar para clasificar la categoría (Boolean, Literal, Resource) y el tipo de respuesta para cada una de las diferentes ontologías que proponen.

Asimismo, se pretende diseñar y desarrollar un módulo para la segunda tarea de predicción de relaciones usando tanto la ontología de DBpedia como la de Wikidata. Al igual que en la tarea uno, se provee de corpus para trabajar.

3. Resultados Obtenidos y Trabajos Futuros

En una primera instancia se ha realizado el relevamiento y análisis de las diferentes estrategias y características empleadas en los sistemas de búsquedas de respuestas semánticos, entre otras herramientas consideradas de utilidad. A partir de este análisis, nos encontramos en la fase de diseño e implementación de los módulos correspondientes a nuestro sistema de búsqueda de respuesta semántico, tales como aquellos para la clasificación de la pregunta y los encargados de la búsqueda y análisis de las posibles respuestas. Esto es un primer paso a nuestro objetivo de implementar nuestro primer prototipo dentro del marco del proyecto de investigación.

La finalidad de nuestra propuesta es la implementación de un Sistema de Búsqueda de Respuesta aplicado sobre la Web Semántica, que nos brinde una respuesta precisa a una pregunta planteada en lenguaje natural, en lugar de una lista de enlaces a documentos como lo hacen los motores de búsquedas tradicionales. Además, como mencionamos en [2], se pretende ampliar

¹<https://smart-task.github.io/2021/>

²<https://smart-task.github.io/2020/>
implícitas en el texto, entre otras.

el conocimiento en el área de Ciencias de la Computación, y por sobre todo en los campos de la Web Semántica y la Ontología.

4. Formación de Recursos Humanos

Durante la realización de esta investigación se espera lograr, como mínimo, la culminación de 2 tesis de grado dirigidas y/o codirigidas por los integrantes del proyecto. Uno de los autores de este trabajo posee una Beca de Estímulo a las Vocaciones Científicas (CIN) 2021.

Finalmente, es constante la búsqueda hacia la consolidación como investigadores de los miembros más recientes del grupo.

Referencias

- [1] Riyanka Manna, Dipankar Das, and Alexander Gelbukh. Question classification in a question answering system on cooking. In *Mexican International Conference on Artificial Intelligence*, pages 103–108. Springer, 2020.
- [2] Matías Oyarzun and Sandra Roger. Tecnologías de sistemas de QA aplicadas a la Web Semántica. In *XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021, Chilecito, La Rioja)*, 2021.
- [3] Nandana Mihindukulasooriya, Mohnish Dubey, Alfio Gliozzo, Jens Lehmann, Axel-Cyrille Ngonga Ngomo, and Ricardo Usbeck. SeMantic Answer Type prediction task (SMART) at ISWC 2020 Semantic Web Challenge. *CoRR/arXiv*, abs/2012.00555, 2020.
- [1] Riyanka Manna, Dipankar Das, and Alexander Gelbukh. Question classification in

Herramientas Informáticas de Dominio Específico para el Desarrollo de Servicios Digitales Innovadores para Comunidades Urbanas y Rurales en el Marco de Ciudades y Regiones Inteligentes

Luis Vivas¹, Mauro Cambarieri¹, Nicolás García Martínez¹, Horacio Muñoz Abbate¹, Marcelo Petroff¹, Hector Ruiz¹

¹ Universidad Nacional de Río Negro. Sede Atlántica
Laboratorio de Informática Aplicada
{lvivas, mcambarieri, ngarciam, hmunoz, mpetroff, hruiz}@unrn.edu.ar

RESUMEN

El objetivo del proyecto es investigar sobre el desarrollo y uso de recursos informáticos de dominio específico para comunidades urbanas y rurales que permitan el desarrollo rápido de servicios públicos en el contexto del desarrollo de ciudades y regiones inteligentes. En base a nuevos paradigmas y tecnologías digitales disponibles - como, por ejemplo: la multi-nube, internet de las cosas, analítica y minería de datos, inteligencia artificial, realidad virtual y realidad aumentada, las aplicaciones móviles, blockchain, entre otras - el proyecto se enfocará en desarrollar activos informáticos para la prestación de servicios digitales innovadores que resulten inteligentes al contexto. El resultado esperado es la apropiación de dichos activos por parte de los ciudadanos y los gobiernos provinciales y municipales y para ello se trabajará en: a) especificar activos informáticos de dominio específico; b) especificar servicios digitales públicos en tiempo de diseño y de ejecución para su rápido desarrollo y despliegue; c) especificar metodologías y buenas prácticas para la integración de dichos activos y servicios; d) desarrollar normas para la utilización segura de los recursos; y e) desarrollar tableros de comando para la toma de decisiones de gobierno y la definición de políticas públicas relacionadas con los servicios en el contexto del desarrollo local y regional.

CONTEXTO

El proyecto comenzará con una revisión del estado del arte, basándose en los resultados obtenidos en el PI 40-C-551 “Herramientas informáticas para el desarrollo de servicios digitales innovadores para comunidades urbanas y rurales en el marco de Ciudades y Regiones Inteligentes (HI-DeSeDi)” poniendo especial énfasis en la transformación digital en el estado como dominio específico y la continuidad

en el diseño y desarrollo de servicios públicos digitales innovadoras en base a: publicaciones científicas, casos de estudio implementados por gobiernos en distintas partes del mundo, y recomendaciones de políticas emitidas por el gobierno nacional, provincial, local, y organismos internacionales. De esta manera, se pondrá foco en las necesidades de los gobiernos municipales y provinciales. En base a los resultados que se obtengan, se desarrollarán activos informáticos reutilizables para el rápido desarrollo de servicios específicos, teniendo en cuenta las últimas tecnologías identificadas en el estado del arte y las características de los gobiernos. Se realizarán estudios que permitirán comparar y evaluar los recursos desarrollados, la generalización de resultados, y la definición de enfoques para su transferencia a otros contextos. Durante todo el proyecto, se pondrá foco en diseminar los resultados obtenidos a través de distintos tipos de publicaciones, como en la formación de recursos humanos.

1. INTRODUCCIÓN

Con el fin de buscar herramientas que ayuden a proveer soluciones a los problemas acuciantes que deben resolver, los gobiernos recurren a las tecnologías digitales y desarrollan iniciativas de Transformación Digital. Con nuevas aplicaciones, herramientas y recursos informáticos los gobiernos avanzan en mejorar los servicios que prestan a los habitantes de sus regiones, los esfuerzos en innovar continúan reflejándose en cuatro etapas de evolución: 1) Digitalización – tendiente a la modernización, mejora de la eficiencia interna y acceso a información de gobierno; 2) Transformación – intenta mejorar procesos internos, estructuras y procedimientos de gobierno; 3) Participación – busca transformar las relaciones entre gobierno, ciudadanos, empresas y otros actores no gubernamentales; y 4) Contextualización – las iniciativas apoyan esfuerzos específicos para el desarrollo, es decir para lograr objetivos específicos de políticas públicas y objetivos de desarrollo sostenible. A esta última etapa se la

denomina Gobernanza Electrónica Dirigida por Políticas. No obstante, en la mayoría de los casos a nivel local, muchas iniciativas se solapan en estas etapas y mientras algunos servicios al ciudadano están mayormente desarrollados que otros, abundan casos del uso de tecnologías “desparejas”, sistemas legados que no interactúan correctamente con otros más modernos, dejando postergadas innovaciones que podrían solucionar los problemas que afectan a los habitantes de la región.

Por su parte, como se ha venido observando, la entrega de servicios públicos digitales también presenta innovaciones. Trabajos relacionados muestran un marco de referencia explicando cómo los sistemas de entrega de estos servicios proveen progresivamente servicios informacionales, mejorados, transaccionales, e integrados, y gradualmente prueban los límites de la innovación incluyendo servicios que promueven la transparencia, son participativos, anticipatorios, personalizados, sensibles e inteligentes al contexto.

Este proyecto se enmarca en los procesos de transformación digital con alto impacto en la resolución de problemáticas gubernamentales a nivel municipal y provincial – en particular, continuar en el desarrollo de un conjunto de recursos emergentes para la rápida entrega de servicios digitales innovadores que resuelvan necesidades de comunidades urbanas y rurales en el contexto de ciudades y regiones inteligentes. Ejemplos de problemas a abordar en el desarrollo de servicios digitales urbanos y rurales son: 1) facilitar la interacción gobierno-ciudadanos, gobierno-empresas y gobierno-gobierno; 2) provisión de servicios turísticos, culturales y sociales sensibles e inteligentes en el actual contexto de pandemia y sus consecuencias; 3) facilitar la integración de personas con discapacidades físicas a través de servicios diseñados con interfaces adaptativas; 4) facilitar el acceso a información para análisis de grandes volúmenes de datos y 5) la adopción de tecnologías integradas al agro y la ganadería para hacer más eficiente el uso y el cuidado de los recursos naturales

Los gobiernos y la población mundial enfrentan grandes desafíos para el desarrollo, a menudo causando pérdida de equilibrio en relaciones y estructuras de poder sociales y económicas; a modo de ejemplo: 1) altas tasas de urbanización y problemas asociados – se prevé que la población urbana mundial crecerá 63% del 2014 al 2050, actualmente 828 millones de personas viven en barrios pobres y carecen de servicios básicos como agua potable y saneamiento (UNDESA, 2014); 2) la falta de empleo estable – del 84 por ciento de la fuerza laboral mundial, el 75% de los trabajadores están empleados con contratos de corto plazo, en empleos informales, o no remunerados (ILO, 2015); y 3) la inequitativa distribución de la riqueza – el 1% de la población posee la mitad de la riqueza mundial

(Credit Suisse Global, 2015).

Considerando los graves desafíos de desarrollo, los países miembros de las Naciones Unidas han adoptado una agenda de desarrollo sostenible que propone 17 objetivos (ODS) a alcanzar en el año 2030. Los ODS consideran particularmente las mejoras a la calidad de vida en diferentes áreas: vida saludable y bienestar (ODS 3); educación y aprendizaje (ODS 4); género (ODS 5); servicios básicos, como agua, saneamiento y electricidad (ODS 6 y ODS 7); empleo y trabajo (ODS 8); conseguir que las ciudades y que los asentamientos humanos sean inclusivos (ODS11); y promover sociedades pacíficas e inclusivas para el desarrollo sostenible, y crear instituciones eficaces, responsables e inclusivas a todos los niveles (ODS 16).

Los gobiernos en distintas partes del mundo recurren a nuevas tecnologías digitales a fin de proveer soluciones a los problemas acuciantes que deben resolver. (Janowski, 2015) afirma que el área de Gobierno Digital cambia constantemente para reflejar los esfuerzos en innovación a través de soluciones tecnológicas que realizan los gobiernos para dar respuesta a las presiones sociales, económicas, políticas y otras, y a las transformaciones que en ellos mismos se producen durante estos procesos.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

El objetivo general del proyecto es investigar sobre las tecnologías digitales y los recursos informáticos de dominio específico emergentes que permitan la rápida entrega de servicios digitales innovadores que puedan satisfacer las necesidades de los habitantes en el contexto del desarrollo de ciudades y regiones inteligentes. Encuadrado en este marco, el proyecto busca desarrollar un conjunto de recursos y activos informáticos, como la definición de mejores prácticas para que los gobiernos municipales y provinciales puedan innovar rápidamente en la prestación de sus servicios. En base a esta premisa, se definen los siguientes objetivos específicos;

O1) Investigar el estado del arte en nuevas tecnologías informáticas para la entrega de servicios digitales de dominio específico.

O2) Desarrollar herramientas, metodologías y mejores prácticas para hacer más eficientes el diseño y uso de recursos informáticos para la rápida entrega de servicios innovadores que satisfagan las necesidades de comunidades urbanas y rurales en el contexto del desarrollo de ciudades y regiones inteligentes en la Provincia de Rio Negro

O3) Proponer un enfoque que permita

generalizar el uso de los recursos de dominio específico para la rápida transformación digital en el ámbito municipal y provincial, en base a la comparación del estado del arte, tal lo citado en el Objetivo Específico 1 (O1), y las herramientas, metodologías y mejores prácticas a desarrollar, tal lo citado en el Objetivo Específico 2 (O2).

3, RESULTADOS OBTENIDOS / ESPERADOS

El equipo que integra el proyecto de investigación pertenece al Laboratorio de Informática Aplicada de la Universidad Nacional de Río Negro, que es una Unidad Ejecutora de Investigación ubicada en el Campus de la Sede Atlántica.

El equipo alcanzó logros en diversos proyectos de I+D realizados por el Laboratorio de Informática Aplicada (LIA). A la fecha son de destacar:

El PI-UNRN 026/09 Conceptos y Modelos de Referencia para la Formulación de una Agenda de Gobierno Electrónico y del Estudio.

El PI-UNRN 40-C-270 Evaluación de Tecnologías de la Información y la Comunicación para el Desarrollo de Ciudades Inteligentes en Río Negro.

El PI-UNRN 40-C-551 Herramientas Informáticas para el Desarrollo de Servicios Digitales Innovadores para Comunidades Urbanas y Rurales en el marco de Ciudades y Regiones Inteligentes (HI-DeSeDi).

Los integrantes del equipo del proyecto han desarrollado varios trabajos de transferencia de tecnología que redundaron en aplicaciones en uso por distintos organismos públicos, como el Sistema de Voto Nominal Legislativo, instalado en la Legislatura de Río Negro; la Plataforma de Abogados del Estado implementado en la Fiscalía de Estado de la Provincia de Río Negro; la App "Viedma Comunidad Digital" en uso por la Municipalidad de Viedma para la zonificación, aforo y control de espacios públicos, en el contexto de pandemia.

Sus integrantes han obtenido en tres oportunidades el Premio Nacional de Gobierno Electrónico en la categoría Proyectos que entrega la SADIO (Sociedad Argentina de Informática) en el marco del Simposio de Informática en el Estado. Los trabajos galardonados fueron:

Desarrollo de una aplicación móvil de intermediación laboral para sectores excluidos del mercado de trabajo. Un instrumento para la definición de políticas públicas. Septiembre 2017

Monitoreo de riesgos de activos de información en la Universidad Nacional de Río Negro. Septiembre 2019

"Aplicación de metodologías y enfoques para el desarrollo de software de dominio específico en el contexto de Gobierno Digital". Octubre 2020

4. 4. FORMACION DE RECURSOS HUMANOS

El grupo de trabajo se encuentra formado por cuatro investigadores formados en la temática, dos investigadores en formación y cinco alumnos avanzados de la carrera Licenciatura en Sistemas. En su marco se desarrolla dos tesis de Maestría en Ciencias de la Computación y se producirán cinco trabajos finales de carrera de grado.

5. BIBLIOGRAFIA

Alarabiat, A., Sa Soares, D., & Estevez, E. (2016). Electronic Participation with a Special Reference to Social Media - A Literature Review. In *Electronic Participation* (pp. 41–52). Springer.

Amugongo, L.M., Nggada, S.N., Sieck, J. (2016). Leveraging on open data to solve city challenges: A case study of Windhoek municipality. In *2016 3rd MEC International Conference on Big Data and Smart City, ICBDS 2016* (pp. 129–134).

Anttiroik, A.-V., Valkama, P., & Bailey, S. J. (2014). Smart Cities in the New Service Economy: Building Platforms for Smart Services. In *AI & SOCIETY2* (pp. 323–334).

Apostol, E., Leordeanu, C., Mocanu, M., & Cristea, V. (2015). Towards a hybrid local-cloud framework for smart farms. In *2015 20th International Conference on Control Systems and Computer Science, CSCS 2015* (pp. 820–824).

Bediroglu, S., Yildirim, V., Nisanci, R. (2016). Building spatial cloud-based local government services. *Proceedings of the Institution of Civil Engineers: Municipal Engineer*, 1(169), 47–60.

Bertot, J. C., Estevez, E., & Janowski, T. (2016a). Digital Public Service Innovation: Framework, Cases, Trends. In *World Public Sector Report 2016: Innovating Public Service Delivery for Sustainable Development*. United Nations Department of Economic and Social Affairs.

Bertot, J. C., Estevez, E., & Janowski, T. (2016b). Universal and contextualized public services: Digital public service innovation framework. *Government Information Quarterly*, 33, 211–222.

<https://doi.org/10.1016/j.giq.2016.05.004>

Christos, G., Xenia, Z., Antonis, R., K, P., Jain, A., Gangadharan, G. R., & Yehiia, T. (2013). Applying Open Innovation Strategies to eGovernment for Better Public Services. In *E-Government Success Factors and Measures: Theories, Concepts, and Methodologies* (pp. 308–330). IGI Global.

Cledou, G., & Barbosa, L. S. (2016). An Ontology for Licensing Public Transport Services. In *9th International Conference on Theory and Practice of Electronic Governance (ICEGOV '15-16)* (pp. 230–239). CONF, New York, NY, USA: ACM. <https://doi.org/10.1145/2910019.2910101>

Clements P. and L. Northrop. *Software Product Lines: Practices and Patterns*. Addison-Wesley Professional, August 2001.

Credit Suisse Global. (2015). *The Global Wealth Report 2015*.

D, K., Levine, D., & Wall, T. (2008). New Development: What are the Challenges in Transferring Lean Thinking to Government? *Public Money & Management*, 28(1).

Dermentzi, E., Tambouris, E., Tarabanis, K. (2016). Cloud computing in eGovernment: Proposing a conceptual stage model. *International Journal of Electronic Government Research*, 12(1), 50–68.

Douwe, V., Estevez, E., & Janowski, T. (2009). Software Infrastructure for e-Government – e-Appointment Service. In *U- and E-Service, Science and Technology* (pp. 141–152). Springer.

Esmaeili, L., & Hashemi G, S. A. (2015). Rural intelligent public transportation system design: Applying the design for re-engineering of transportation eCommerce system in Iran. *International Journal of Information Technologies and Systems Approach*, 8(1), 1–27.

Estevez, E., & Janowski, T. (2013). Electronic Governance for Sustainable Development — Conceptual framework and state of research. *Government Information Quarterly*, 30, S94–S109. <https://doi.org/10.1016/j.giq.2012.11.001>

Estevez, E., Janowski, T., & Dzhupova, Z. (2013). Electronic Governance for Sustainable Development – How EGOV Solutions Contribute to SD Goals?, 92–101.

Estevez, E., Lopes, N., & Janowski, T. (2015). *Smart Cities for Sustainable Development - Reconnaissance Study*.

Evans E., *Domain-Driven Design: Tackling Complexity in the Heart of Software*. Addison-Wesley, 2003

Finkelievich, S. (2014). Innovación abierta en la sociedad del conocimiento redes transnacionales y

comunidades locales. Instituto de Investigaciones Gino Germani.

Gonzalez-Huerta J, S. Abrahão, E. Insfran: Un enfoque Multi-modelo para la Introducción de Atributos de Calidad en el Desarrollo de Líneas de Producto Software. XVI Jornadas de Ingeniería del Software y Bases de Datos. Septiembre de 2011

Guglielmi, G. J. (2016). Open data y servicio público. Los datos públicos abiertos son un servicio público. *Revista General de Derecho Administrativo*, 2016(41), 1–14.

ILO. (2015). *World Employment Social Outlook: The changing nature of jobs*. Retrieved from http://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_368626.pdf

ITU. (2015). *ICT Facts & Figures: The World in 2015*. Retrieved from <http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf>

Janowski, T. (2015). Digital government evolution: From transformation to contextualization. *Government Information Quarterly*, 32(3), 221–236. <https://doi.org/10.1016/j.giq.2015.07.001>

Janowski, T., Ojo, A., & Estevez, E. (2007). Rapid Development of Electronic Public Services – Software Infrastructure and Software Process. *Proceedings of the 8th Annual International Conference on Digital Government Research: Bridging Disciplines & Domains*, 294–295.

Janssen, M., & Estevez, E. (2013). Lean Government and Platform-based Governance - Doing More with Less. *Government Information Quarterly*, 30(1), S1–S8. <https://doi.org/10.1016/j.giq.2012.11.003>

Janssen, M., & Zuiderwijk, A. (2014). Infomediary Business Models for Connecting Open Data Providers and Users. *Social Science Computer Review*, 32(5), 694–711.

Koussouris, S., Kokkinakos, P., Panopoulos, D., Askounis, D., Ramfos, A., Georgousopoulos, C., & Wittern, E. (2012). Citizens Collaboration and Co-Creation in Public Service Delivery: The COCKPIT Project. *International Journal of Electronic Government Research*, 8(3), 33–62.

Krueger, C W: Homeaway's transition to software product line practice: Engineering and business results in 60 days. In *Proc. of the 12th International Software Product Line Conference (SPLC'08)*, Ireland, pages 297–306. IEEE, September 2008

Lopez, V., Minana, G., Sanchez, O., Gonzalez, B., Valverde, G., & Caro, R. (2015). Big+Open Data: Some applications for a Smartcity. In *3rd IEEE International Conference on Progress in Informatics and Computing, PIC 2015* (pp. 384–389).

- Maccani, G., Donnellan, B., & Helfert, M. (2015). Open Data Diffusion for Service Innovation: An Inductive Case Study on Cultural Open Data Services.
- Macintosh, A., Coleman, S., & Schneeberger, A. (2009). eParticipation: The Research Gaps. In Proceedings of ePart 2009. Springer Verlag.
- Mc Dermott, P. (2010). Building Open Government. *Government Information Quarterly*, 27(4), 401–413.
- Mergel, I., & Desouza, K. (2013). Implementing Open Innovation in the Public Sector: The Case of Challenge.gov. *Public Administration Review*, 73(6), 882–890.
- Nafis, F., Yousfi, S., Chiadmi, D. (2016). How big open data can improve public services. *Lecture Notes in Electrical Engineering*, 381, 607–612.
- Ojo, A., Janowski, T., & Estevez, E. (2007). A Composite Domain Framework for Developing Electronic Public Services. In Proceedings of the International Conference on Software Engineering Theory and Practice (SETP 2007), Orlando, Florida, USA, July 2007.
- Oquendo, F.: p-Method: A Model-Driven Formal Method for Architecture-Centric Software Engineering. *ACM SIGSOFT Software Engineering Notes*. Volume 31 Number 3. (2006).
- Domain Driven Design Reference. Definitions and Pattern Summaries. Eric Evans.2015
- Panopoulou, E., Tambouris, E., & Tarabanis, K. (2010). EParticipation initiatives in Europe: Learning from practitioners. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6229 LNCS, 54–65. https://doi.org/10.1007/978-3-642-15158-3_5
- Pasini, A., Estevez, E., Pesado, P., & Boracchia, M. (2016). Una metodología para evaluar la madurez de servicios universitarios. In XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016) (pp. 636–646).
- Paskaleva, K. A. (2011). The Smart City: A Nexus for Open Innovation? *Intelligent Buildings International*, 3(3), 153–171.
- Quarati, A., Clematis, A., D’Agostino, D. (2016). Delivering cloud services with QoS requirements: Business opportunities, architectural solutions and energy-saving aspects. *Future Generation Computer Systems*, 55(1), 403–427.
- Saxena, S., Kumar Sharma, S. (2016). Integrating Big Data in “e-Oman”: opportunities and challenges. *Info*, 18(5), 79–97.
- Tayab, M., Zhou, W., Zhao, M., Li, S. (2016). Big data and public services for environmental monitoring system. In ICCSE 2016 - 11th International Conference on Computer Science and Education (pp. 139–143).
- Traunmuller, R. (2014). Open Government and Electronic Government: Some Considerations. In 3rd International Conference on Electronic Government and the Information Systems Perspective, EGOVIS 2014 (pp. 201–207).
- UNDESA. (2014). World Urbanization Prospects, the 2014 Revision. (United Nations Department of Economic and Social Affairs, Ed.). <https://doi.org/10.4054/DemRes.2005.12.9>
- Wahsh, M.A., Dhillon, J. S. (2015). An investigation of factors affecting the adoption of cloud computing for E-government implementation. In 2015 IEEE Student Conference on Research and Development, SCOREd 2015 (pp. 323–328).
- Yahaya, J.H., Deraman, A., Abai, N.H.Z., Mansor, Z., Jusoh, Y. Y. (2016). Business intelligence and big data analytics for organizational performance management in public sector: The conceptual framework. *Advanced Science Letters*, 22(8), 1919–1923.
- Zhang, H., & Liu, X.-Y. (2016). Research on the construction of intelligent agriculture service system based on “internet+” in Hunan province. In Conference Proceedings of the 4th International Symposium on Project Management, ISPM 2016 (pp. 576–579).

Coordinación de dispositivos de IoT para resolver problemas específicos de la distribución de energía eléctrica

Oscar Testa¹; Rubén Pizarro¹; Darío Segovia¹

¹ Facultad de Ciencias Exactas y Naturales
Universidad Nacional de La Pampa
Av. Uruguay 151 – (6300) Santa Rosa – La Pampa – Argentina
Tel.: +54-2954-425166– Int. 28
[otesta, ruben]@exactas.unlpam.edu.ar

Resumen

Este proyecto de investigación, deriva del proyecto “Ingeniería de Software: Composición de servicios en ambientes ubicuos”, desarrollado también en el ámbito de la Facultad de Ciencias Exactas y Naturales de la UNLPam.

Como resultado de dicho proyecto de Investigación, se logró implementar un framework para la coordinación de dispositivos ubicuos, a través de la utilización de los estándares de SOA para servicios. Como parte de las líneas futuras de investigación de dicho proyecto, surgió la necesidad de realizar implementaciones en ambientes reales, donde se pueda poner a prueba dicho entorno.

Se plantea entonces, para este proyecto, la posibilidad de implementar el framework para un caso de estudio específico y real, una distribuidora eléctrica. El caso de estudio es el de poder monitorear a través de la utilización de sensores, cámaras y dispositivos distintas maquinarias que forman parte del servicio eléctrico (transformadores, cámaras o subestaciones) con el fin de prevenir accidentes que puedan dañar tanto los materiales como el servicio que brinda la empresa hacia la comunidad.

Se espera como resultado de este proyecto, obtener una implementación real del framework de coordinación de dispositivos, el cual abre las

puertas para una producción de este tipo de soluciones para todo tipo de ambientes.

Palabras clave: sistemas ubicuos, SOA, servicios, composición de servicios, coreografías.

Contexto

El presente trabajo se enmarca en el Proyecto de Investigación: Coordinación de dispositivos de IoT para resolver problemas específicos de la distribución de energía eléctrica – Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa. Las líneas aquí presentadas actualmente forman parte de las bases de un anteproyecto de tesis de maestría en ingeniería de software por la Universidad Nacional de San Luis.

Introducción

En la actualidad, las actividades cotidianas del hombre se han hecho dependientes de una gran cantidad de dispositivos electrónicos tales como: ordenadores personales, ordenadores portátiles, teléfonos móviles, PDAs, tabletas, sensores de muchas y diversas utilidades, entre otros; los cuales logran comunicarse entre sí gracias a diversos protocolos de comunicación inalámbrica, redes de celulares, redes de área local (LAN), redes de área extensa (WAN), Bluetooth, etc. Estamos en la presencia de nuevos dispositivos de comunicación, lo que conlleva un nuevo escenario social, donde la interacción permanente con estos elementos es ineludible. Por ejemplo, la cantidad de móviles

existentes en el mercado se aproxima a la cantidad de habitantes mundiales, según un informe de la Unión Internacional de Comunicaciones. En este informe, se estima que hasta finales del año 2014 hubo casi 7.000 millones de suscripciones de telefonía celular, lo que corresponde al 96 % de la población global; es decir, la cantidad de usuarios de telefonía móvil se acerca al número de personas que viven en el planeta [1].

Los avances de las comunicaciones entre dispositivos ha permitido que estos sean generadores y consumidores de servicios al mismo tiempo, es decir, de acuerdo a las capacidades del dispositivo puede no solo obtener, sino también ofrecer a otros equipos sus funciones y así cooperar entre ellos. La tendencia actual es hacia los ambientes ubicuos, los cuales se caracterizan por estar poblados de numerosos dispositivos que, gracias a la integración extrema de los elementos electrónicos, son invisibles al usuario y están en permanente rastreo de la actividad humana [2].

Dispositivos ubicuos son todos aquellos dispositivos que pueden existir en todas partes, es decir, son dispositivos electrónicos que tienen capacidad de procesamiento y comunicación y pueden ser encontrados en lugares diversos de la vida cotidiana.

La computación ubicua es un desarrollo tecnológico que intenta que las computadoras no se perciban en el entorno como objetos diferenciados, y que la utilización por parte de los seres humanos sea lo más transparente y cómoda posible, facilitando de esta manera la integración en la vida cotidiana. Desde hace varios años los dispositivos ubicuos han ganado importancia y presencia en la vida cotidiana de las personas, debido principalmente a que: poseen distintos tipos de sensores (posicionamiento, proximidad, luminosidad, temperatura, etc.), facilitan la conectividad incluso en áreas con poca señal o acceso a las redes, permiten la convergencia tecnológica (computo, medios, telefonía, etc) y brindan acceso a servicios de distinta índole (mapas, ayudas, etc).

Por composición entendemos la forma en que se pueden combinar o enlazar un número indeterminado de dispositivos para llevar adelante una tarea determinada. En ambientes ubicuos, la composición de dispositivos, presenta nuevos desafíos tales como: la heterogeneidad (ya sea por la diversidad de dispositivos involucrados, como por la presencia de dispositivos de varios fabricantes), las contingencias de los dispositivos y la personalización de los mismos (por ej. provisión de servicios de acuerdo a las preferencias del usuario). Dado que los dispositivos en donde los servicios son ejecutados poseen limitaciones de recursos (ej. poca memoria y batería), se deben hacer consideraciones especiales respecto a la eficiencia y rendimiento de la composición de servicios [3].

La composición en este tipo de ambientes implica que los dispositivos deben dialogar entre ellos para poder compartir los servicios que ofrecen con la finalidad de obtener un servicio con valor agregado, o bien para abordar la solución de una problemática particular, como podría ser la seguridad de un hogar, o la seguridad vial, por mencionar algunos ejemplos.

Si bien hoy en día podemos decir que distintos sensores o dispositivos se pueden comunicar entre ellos, compartiendo de alguna manera sus servicios, generalmente lo realizan a partir de protocolos propietarios y sin seguir definiciones estándares, provocando que otros componentes de otros proveedores (o incluso de los mismos) no puedan ser utilizados. Esto obviamente representa una importante limitación en la composición de dispositivos ubicuos. Adicionalmente la composición de dispositivos ubicuos presenta un nuevo desafío. Los mecanismos de composición en ambientes masivos, necesitan hacer frente a las distintas contingencias que pueden ocurrir con estos dispositivos. Los dispositivos ubicuos tienen distintas limitantes como son la cantidad de memoria disponible, la durabilidad de la batería, la disponibilidad de acuerdo a la red del

lugar donde se encuentre en un momento determinado. Todas estas variantes hacen que la composición de dispositivos¹ ubicuos se transforme en un área de investigación muy importante donde los avances no han sido claros al día de hoy[3].

Líneas de Investigación y Desarrollo

La computación orientada a servicios, y en particular los servicios web en ambiente de internet, proporcionan mecanismos para la composición de servicios. Dichos mecanismos, como las orquestaciones, son aspectos bien conocidos de la computación orientada a servicios que permiten construir sistemas de negocio complejos y aplicaciones a partir de una gran cantidad de servicios heterogéneos, simples y distribuidos. Podría pensarse que son aplicables a ambientes ubicuos. Sin embargo, en contextos como puede ser la Internet de las Cosas (IoT) donde los servicios son dinámicos, móviles, menos fiables y dependientes del dispositivo, los mecanismos de composición establecidos para servicios web no es directamente aplicable [8].

Adicionalmente la composición de múltiples dispositivos ubicuos presenta nuevos desafíos que no son compatibles con la composición de servicios web. En particular, los mecanismos de composición en ambientes masivos como lo es el de dispositivos móviles, necesita hacer frente las distintas contingencias que pueden ocurrir con estos elementos, así como también contemplar la heterogeneidad de los mismos.

Estos dispositivos tienen distintas limitantes como son la cantidad de memoria disponible, la durabilidad de la batería, la disponibilidad de acuerdo a la red del lugar donde se encuentre en

un momento determinado. En ambientes ubicuos, la disponibilidad y confiabilidad de los dispositivos no puede ser garantizada. Todas estas dificultades hacen que la composición de dispositivos se transforme en un área de investigación muy importante donde los avances no han sido claros al día de hoy[3].

Finalmente existen distintos proyectos en la actualidad donde se intenta integrar sensores y dispositivos ubicuos a la vida cotidiana. Específicamente podemos mencionar la domótica, donde varios dispositivos y sensores deben actuar en coordinación para prevenir un incidente de seguridad (ya sea por robo o por incendio) en nuestros hogares. Sin embargo, existen áreas de aplicación más relevantes.

En la industria, existe lo que se llama Industria 4.0 [9], donde lo que se intenta es integrar dentro de una planta fabril la intercomunicación de todos los dispositivos que componen la cadena de producción con el fin de que coordinen entre ellos las tareas a realizar en base a los tiempos a cumplir, stocks disponibles, demanda en línea de los productos, etc. Otra área donde los dispositivos ubicuos están ganando importancia es la automotriz, donde los esfuerzos se enfocan en que distintos sensores monitoreen funciones vitales del conductor (como es el caso de presión arterial, pulsaciones, etc) y en caso de que detecten anomalías actúen en conjunto con otros dispositivos del vehículo para evitar accidentes.

En este sentido, y en esta investigación en particular es que abordamos la implementación de un framework de coreografía de servicios con dispositivos ubicuos, la cual está definida por Testa, et al. en[11]. A partir de la utilización de este framework es que se propone implementarlo sobre un sistema real de una distribuidora eléctrica, principalmente para el monitoreo y automatización de subestaciones transformadoras bajo nivel. Se decidió trabajar sobre este tipo de elementos en parte por la necesidad de la distribuidora y en parte por la complejidad media del proyecto.

Resultados y Objetivos

¹ Si bien los autores se refieren a la composición de servicios, se hace dentro de un contexto de dispositivos ubicuos, lo cual a los fines de este trabajo se puede interpretar como composición de dispositivos, haciendo que la terminología para este caso particular sea más adecuada.

La proliferación de dispositivos de computación ubicuos e interconectados (PDAs, tabletas, móviles, etc), así como los recientes avances en la tecnología de radio frecuencia y las redes de sensores están fomentando la creación de ambientes donde las aplicaciones de internet y los servicios se están haciendo mas populares y necesarias para los usuarios de móviles. La composición de servicios a través de múltiples dispositivos móviles presenta un nuevo desafío el cual no es compatible con la composición de servicios como se plantea actualmente. En particular, los mecanismos de composición en ambientes masivos como lo es el de dispositivos móviles, necesita hacer frente las distintas contingencias que pueden ocurrir con estos elementos, así como también contemplar la heterogeneidad de los mismos. Estos dispositivos tienen distintas limitantes como son la cantidad de memoria disponible, la durabilidad de la batería, la disponibilidad de acuerdo a la red del lugar donde se encuentre en un momento determinado. Todas estas variantes hacen que la composición de servicios incluyendo dispositivos móviles se transforme en un área de investigación muy importante donde los avances no han sido claros al día de hoy [3].

En este tipo de ambientes, mecanismos automáticos y dinámicos son necesarios para la composición de servicios, ya que de esta forma se puede compensar la falta de disponibilidad de un servicio en un momento determinado [8].

Se plantea entonces que hoy en día la composición de servicios es aun mas compleja debido a la incorporación de estos dispositivos que se mencionan en el párrafo precedente. La composición tiene una complejidad mayor aun cuando se trata de servicios dependientes y especialmente en ambientes críticos.

Si bien a este problema existente se encontró solución, a través de la implementación de un framework de ejecución de coreografías en ambientes ubicuos, tal como se plantea en [11], se necesita hacer pruebas y más investigaciones a partir de una implementación real.

Por lo antes expuesto se deducen las siguientes preguntas de investigación:

- ¿El framework de ejecución de coreografías propuesto, puede implementarse para una situación real?
- ¿Qué mejoras son necesarias aplicar sobre el framework para que el mismo pueda ser aplicado a ambientes reales de ejecución?
- ¿Qué tipo de hardware se necesita para una implementación real, o alcanza con utilizar placas académicas de uso general como Arduino?
- ¿Los protocolos de comunicación seleccionados, son los correctos, o debe ampliarse la implementación y utilización de protocolos?

Por lo tanto, el proyecto plantea los siguientes objetivos:

- Realizar una implementación real del framework de ejecución de coreografías en ambientes ubicuos, en un caso de estudio particular y real.
- Realizar las mejoras y ajustes necesarios al framework para que pueda ser utilizado comercialmente.
- Poner a prueba la validez de la utilización de los protocolos de comunicación establecidos por SOA.
- Difundir los resultados obtenidos con publicaciones de los resultados parciales y finales.

La hipótesis principal es poner a prueba si los mecanismos principales y fundacionales de SOA son aplicables para la composición de coreografías en ambientes ubicuos, en un ambiente real. Se tienen fuertes indicios, por los estudios e investigaciones llevadas adelante en el proyecto de Investigación sobre el cual se basa el presente trabajo, que los mismos serán suficientes y que serán de utilidad en este tipo de ambientes.

Se espera, además logros en términos de formación en recursos humanos:

Darío Segovia, se espera que los resultados de este trabajo se vean reflejados en la Tesis de Maestría.

Se espera además, la generación de informes de investigación y conferencias con el objeto de difundir lo actuado. De la misma manera, se espera generar más aplicaciones en el campo de la ingeniería de software relacionadas con aspectos específicos del framework de composición propuesto y puesto a consideración en este trabajo.

Formación de Recursos Humanos

Además de los resultados esperados en el punto 3, se espera como resultado en la formación de recursos humanos, la continuación de esta misma línea de proyecto como tesis de maestría y doctoral de alguno(s) de los investigadores. También se espera lograr una mayor interrelación con la Universidad Politécnica de Madrid para la aplicación de las metodologías aquí presentadas en los proyectos de Ingeniería de Software Empírica. Adicionalmente, se espera que otras tesis de Maestría, así como tesinas de Licenciatura surjan a partir de los logros obtenidos en la presente línea de investigación.

Bibliografía

- [1] U. I. d. T. (UIT), “Unión Internacional de Telecomunicaciones.” <https://www.itu.int/net/pressoffice/pressreleases/2014/23-es.aspx>, 10 2015.
- [2] M. Weiser, “Hot topics-ubiquitous computing,” *Computer*, vol. 26, pp. 71–72, Oct 1993.
- [3] Q. Z. Sheng, X. Qiao, A. V. Vasilakos, C. Szabo, S. Bourne, and X. Xu, “Web services composition: A decade’s overview,” *Information Sciences*, vol. 280, no. 0, pp. 218–238, 2014.
- [4] M. Viroli, “On competitive self-composition in pervasive services,” *Science of Computer Programming*, vol. 78, no. 5, pp. 556–568, 2013. Special section: Principles and Practice of Programming in Java 2009/2010 & Special section: Self-Organizing Coordination.
- [5] S. W. Loke, “Supporting ubiquitous sensor-cloudlets and context-cloudlets: Programming compositions of context-aware systems for mobile users,” *Future Generation Computer Systems*, vol. 28, no. 4, pp. 619–632, 2012.
- [6] F. Palmieri, “Scalable service discovery in ubiquitous and pervasive computing architectures: A percolation-driven approach,” *Future Generation Computer Systems*, vol. 29, no. 3, pp. 693–703, 2013. Special Section: Recent Developments in High Performance Computing and Security.
- [7] S. Najar, M. K. Pinheiro, and C. Souveyet, “A New Approach for Service Discovery and Prediction on Pervasive Information System,” *Procedia Computer Science*, vol. 32, pp. 421–428, 2014. The 5th International Conference on Ambient Systems, Networks and Technologies (ANT-2014), the 4th International Conference on Sustainable Energy Information Technology (SEIT-2014).
- [8] G. Cassar, P. Barnaghi, W. Wang, S. De, and K. Moessner, “Composition of services in pervasive environments: A Divide and Conquer approach,” in *Computers and Communications (ISCC)*, 2013 IEEE Symposium on, pp. 000226–000232, July 2013.
- [9] Wikipedia, “Industria 4.0 — Wikipedia, La enciclopedia libre,” 2016. [Internet; descargado 4-noviembre-2016].
- [10] H.-I. Yang, R. Bose, A. (Sumi) Helal, J. Xia, and C. Chang, “Fault-Resilient Pervasive Service Composition,” in *Advanced Intelligent Environments (A. D. Kameas, V. Callagan, H. Hagraas, M. Weber, and W. Minker, eds.)*, pp. 195–223, Springer US, 2009.
- [11] Oscar A. Testa, Efraín R. Fonseca C., Germán Montejano and Oscar Dieste, “Coordination of Ubiquitous Devices in Pervasive Environments: A Proposal Based on WS-CDL” in *38th International Conference of*

the Chilean Computer Science Society. SCCC
2019.

MODELANDO ONTOLOGÍAS CON PATRONES EN UN AMBIENTE GRÁFICO WEB DE INGENIERÍA ONTOLÓGICA

Tomás L. Quiñonez¹ Christian N. Gimenez¹ Laura A. Cecchi¹ Pablo R. Fillottrani^{2,3}

email: tomas.quinonez@est.fi.uncoma.edu.ar,
{christian.gimenez,lcecchi}@fi.uncoma.edu.ar, prf@cs.uns.edu.ar

Grupo de Investigación en Lenguajes e Inteligencia Artificial,
Departamento de Teoría de la Computación - Facultad de Informática
UNIVERSIDAD NACIONAL DEL COMAHUE

Laboratorio de I&D en Ingeniería de Software y Sistemas de Información, Departamento
de Ciencias e Ingeniería de la Computación
UNIVERSIDAD NACIONAL DEL SUR

Comisión de Investigaciones Científicas de la provincia de Buenos Aires (CIC)

RESUMEN

La presente investigación se desarrolla mediante el trabajo colaborativo de docentes investigadores de la Universidad Nacional del Comahue (UNCo) y de la Universidad Nacional del Sur (UNS), en el contexto de proyectos de investigación financiados por las universidades indicadas.

El objetivo general de esta línea de investigación y desarrollo es el diseño e implementación de una metodología que permita recomendar a los usuarios Patrones de Diseño Ontológico para la ontología modelada. Como paso intermedio y a partir de un relevamiento, se construye un catálogo de patrones aceptados por la comunidad. Luego, para los distintos patrones dentro del catálogo, se realiza una comparación de sus componentes con los componentes de la ontología modelada por el usuario, permitiendo la identificación de los mismos en dicha ontología, a través de servicios de razonamiento. El resultado de esta implementación será integrado a crowd, la cual es una herramienta para el modelado visual ontológico, que utiliza lenguajes de modelado conceptual y que es desarrollada por nuestros grupos de investigación.

Palabras Clave: Ingeniería de Software basada en Conocimiento, Patrones de Diseño Ontológico, Ontologías, Lógicas Descriptivas.

CONTEXTO

Este trabajo está parcialmente financiado

por la Universidad Nacional del Comahue, en el marco del proyecto de investigación *Tecnologías semánticas para el desarrollo de agentes inteligentes (04/F020)*, y con una Becas de Investigación en la categoría Graduado de Perfeccionamiento, por la Universidad Nacional del Sur a través del proyecto de investigación *Tecnologías semánticas para acceso e integración de datos (24/N049)* y por el Consejo Interuniversitario Nacional (CIN) con una Beca de Estímulo a las Vocaciones Científicas para Estudiantes 2020. Los proyectos de investigación tienen una duración de cuatro años, la beca de perfeccionamiento una duración de 2 años, finalizando en diciembre de 2022, y la beca CIN una duración de 1 año, finalizando esta última en agosto de 2022.

1. INTRODUCCIÓN

La Ingeniería Ontológica estudia los métodos y metodologías que guían a los modeladores en el diseño, desarrollo, implementación, mantenimiento, uso y publicación de ontologías [10, 15]. Si bien existen diversos enfoques para definir el concepto de ontología [5], en este trabajo utilizaremos el enfoque dado en [9], el cual define una ontología como equivalente a una base de conocimiento en Lógica Descriptiva [1]. Esto agrega una dimensión más a la complejidad del desarrollo, debido a que requiere de un íntimo conocimiento de estos

formalismos.

Actualmente, existen varias herramientas de Ingeniería Ontológicas para el desarrollo de ontologías. Sin embargo, estas herramientas presentan dos falencias claramente identificadas. Si bien las visualizaciones proveen un adecuado nivel de abstracción para ontologías, ningún método de visualización ha sido ampliamente aceptado [3]. Asimismo, recientes revisiones de la literatura sugieren que los ambientes actualmente disponibles presentan un débil soporte para la Ingeniería Ontológica, debido a la fragmentación de las metodologías en diversas herramientas [16].

Una posible forma de asistir al modelador y contribuir en la solución a alguno de estos problemas es el uso de patrones. Los patrones ontológicos pueden ser vistos como pequeñas ontologías bien definidas y aceptadas por la comunidad, usadas como modelos *templates*, para ser incorporados a la ontología en desarrollo. En este sentido, el uso de Patrones de Diseño Ontológico (ODP) [7], junto con paradigmas de modelado gráfico que soporten lenguajes estándar, permitirá que los modeladores reduzcan significativamente el esfuerzo y el costo de desarrollo y mantenimiento de sus ontologías, reusando soluciones formalizadas.

Sin embargo, los enfoques para la Ingeniería Ontológica basados en patrones requieren, por un lado, de la existencia de un conjunto de patrones adecuados y aceptados en la comunidad para ser reusados. Y por el otro, de metodología apropiadas que soporten la elicitación de estos patrones y su aplicación en la construcción de nuevos modelos. Por otra parte, un modelador que desea construir, mantener o validar una ontología a través del reuso de patrones, posee poca asistencia en las herramientas de desarrollo, respecto del uso de estructuras lógicas, que generalmente son poco amigables, haciendo así que las ontologías sean difíciles de comprender.

En este trabajo, se propone diseñar una

metodología para el modelado de ontologías basado en patrones, que sugiera ODP a partir del modelo en desarrollo. Estos patrones se obtendrán de un listado ampliamente aceptado por la comunidad de expertos en el tema¹.

Asimismo, se propone desarrollar una herramienta que implemente dicha metodología, capaz de recomendar patrones que puedan ser aplicados en una ontología. Dicha herramienta consistirá principalmente de una *API REST* la cual podrá ser utilizada, entre otras, por *crowd*, una herramienta visual Web de soporte a la Ingeniería Ontológica, permitiendo la utilización de las funcionalidades de dicha API. Se espera que los modeladores puedan utilizar *crowd* para diseñar sus ontologías, a partir del reuso, extensión e integración de uno o varios patrones, acelerando el proceso de desarrollo y mejorando la calidad de sus ontologías al hacerlas más modulares y reusables.

Este trabajo está estructurado como sigue. En la siguiente sección, se describe la línea de investigación actual. En la Sección 3, se indican algunos resultados obtenidos y trabajos futuros. Finalmente, en la última sección se comentan aspectos referentes a la formación de recursos humanos en esta temática.

2. LÍNEA DE INVESTIGACIÓN Y DESARROLLO

En el proyecto de investigación *Tecnologías Semánticas para el desarrollo de Agentes Inteligentes*, UNCo, se investigan técnicas de representación de conocimiento y razonamiento, metodologías de modelado ontológico y mecanismos para la interoperabilidad tanto a nivel de proceso como de datos. El empleo efectivo de los conceptos y conocimientos adquiridos da soporte a comunidades en el desarrollo de ontologías.

Por otro lado, en el proyecto de investigación *Tecnologías semánticas para acceso e integración de datos*, UNS, se desarrollan metodologías y herramientas que asisten la interoperabilidad semántica de información y de servicios en la Web,

¹Listado de patrones en ontologydesignpatterns.org

privilegiando los últimos avances en el área de lenguajes de representación del conocimiento, ontologías y modelado conceptual.

Ambos proyectos comparten el perfil de investigación de este trabajo, en el que se estudian entre otros, sobre temas afines a las Ontologías, las Lógicas Descriptivas y la Ingeniería de Software basada en Conocimiento. En el presente trabajo de investigación, se hace foco a las tareas de Ingeniería Ontológica, donde el modelador se enfrenta a diversos desafíos a resolver respecto del diseño, mantenimiento y validación de ontologías. Muchos de estos desafíos son recurrentes en el modelado, aun en diferentes dominios, y donde el reuso de soluciones es una forma de atacar estas dificultades.

Una posible técnica de reuso de estas soluciones es a través de Patrones de Diseño Ontológico (ODP) [7]. Los patrones brindan una ayuda al usuario durante la etapa de diseño, al reutilizarlos como guías para determinar si la ontología que está diseñando, está siendo correctamente desarrollada. De esta manera, los ODP son importantes ya que promueven el reuso de soluciones a problemas recurrentes de modelado.

En [13], se presentan un conjunto de patrones agrupados por algunas características de los ODP, como por ejemplo patrones que representan estructuras de datos, sin enfocarse en los dominios. Esto hace que la búsqueda de un patrón en particular sea complicada, aún para desarrolladores expertos en algún dominio. Por otra parte, la mayoría de los patrones se encuentran caracterizados en el lenguaje OWL, lo que hace complejo su uso para usuarios que desconocen dicho lenguaje formal.

En la literatura, existen pocas herramientas desarrolladas para el modelado basado en patrones que además provean soporte gráfico. CoModIDE [12] permite la composición de ontologías basada en una biblioteca de ODP. Sin embargo, utiliza un lenguaje gráfico *ad-hoc* basado en grafos para el modelado, lo que implica familiarizarse con este nuevo lenguaje. ODPReco [17] es una herramienta en desarrollo

de recomendación de patrones, pero no provee soporte gráfico. Asimismo, existen herramientas visuales que soportan lenguajes de modelado conceptual [11, 4], aunque tampoco proveen metodologías para el uso de patrones.

Por consiguiente, las tareas de Ingeniería Ontológica en conjunción con la integración de metodologías y buenas prácticas en el uso de patrones (por ejemplo, que pueda ser seleccionado en forma amigable de un catálogo) en ambientes gráficos de modelado ontológico, es una arista no explorada en profundidad [14, 8]. En consecuencia, se propone acercar ambas dimensiones e implementar en una herramienta Web, una metodología modular para la Ingeniería Ontológica basada en patrones, que utilice lenguajes de modelado conceptual estándar.

Así, esta línea de investigación tiene como objetivo el desarrollo de una metodología para el modelado de ontologías basado en patrones. Dicha metodología recibe como entrada una ontología y sugiere ODP teniendo en cuenta dicho modelo. Para la identificación de los patrones se hará uso de servicio de razonamiento sobre razonadores preexistentes, como Racer y Konclude. Como objetivo intermedio, se plantea la clasificación y construcción de un catálogo digital de ODP para integrar a la metodología propuesta. Finalmente, se implementará un prototipo de dicha metodología que luego se integrará a la herramienta visual Web de Ingeniería Ontológica *crowd* [2].

crowd es una herramienta Web de modelado ontológico, desarrollada por miembros de nuestro grupo de investigación. Su objetivo es proveer una interfaz visual para las tareas asociadas a la Ingeniería Ontológica, utilizando múltiples lenguajes gráficos de modelado, tales como UML, EER y ORM 2 [6]. Además, es capaz de validar la consistencia de los modelos, gracias a que posee razonamiento automático integrado. Actualmente, *crowd* no provee soporte metodológico para el desarrollo de ontología

basado en ODP. La incorporación de dicha metodología a la herramienta es de gran interés, ya que provee una ayuda extra al usuario en el desarrollo de las ontologías, para resolver problemas de diseño recurrentes en diferentes dominios.

3. RESULTADOS OBTENIDOS Y TRABAJO FUTURO

Inicialmente, se realizó un relevamiento sobre los distintos tipos de Patrones de Diseño Ontológico que existen, para poder determinar cuál es el que mejor se adaptaría al trabajo a desarrollar. Los tipos analizados fueron los siguientes: Estructurales, de Correspondencia, de Contenido, de Razonamiento, de Presentación, y Léxico-Sintácticos [11]. Bajo este análisis, se destacaron los Patrones de Contenido debido a la naturaleza de los mismos.

Los Patrones de Contenido son pequeñas ontologías bien definidas y utilizadas como componentes de modelado durante el proceso de Ingeniería Ontológica. Estos patrones proveen soluciones bien probadas a problemas de modelado recurrentes bajo un dominio en particular. Por lo tanto, resulta útil que un usuario que desea desarrollar una ontología, pueda hacer uso de estos patrones y componerlos, de forma tal que la ontología final deseada sea una composición de los mismos. Debido a esto, en esta línea de investigación se decidió trabajar con este tipo de patrones. A partir de esto, se realizó un análisis y selección sobre distintos Patrones de Contenido propuestos por la comunidad de expertos². Estos patrones serán integrados a la metodología y recomendados a los usuarios en caso de existir elementos de dichos patrones en la ontología modelada por él. Para esto, se diseñó un criterio de selección de patrones para poder ser utilizados en este trabajo.

Respecto del desarrollo de la metodología para el diseño de ontologías basado en patrones se han tenido en cuenta las siguientes consideraciones. Un catálogo de Patrones de

Contenido seleccionados en formato *OWL* estarán disponibles. De esta forma, se puede realizar una comparación entre los axiomas de los patrones y los axiomas de la ontología del usuario. La metodología toma como entrada la ontología del usuario en formato *OWL*. Luego, se realizan consultas a un razonador sobre dicha ontología. Una clase de consulta que se realiza es sobre los nombres de clases, relaciones y tipos de datos. Así, se realiza una comparación sobre dichos nombres, con los ya presentes en los patrones. Otra clase de consulta es sobre los tipos de axiomas presentes en la ontología. Con esto, se determina si los distintos tipos de axiomas presentes en los patrones ocurren dentro de la ontología. Para estas dos consultas se hace uso de una estructura de datos auxiliar, que contiene una serie de sinónimos para los nombres de los elementos de todos los patrones presentes en la herramienta. Finalmente, la metodología retorna al usuario una lista con todos los patrones detectados total o parcialmente en la ontología de entrada.

Como trabajo futuro, se pretende realizar la implementación de la metodología en una herramienta como una *API REST*. Esta decisión fue tomada teniendo en cuenta que, la funcionalidad de la misma no debería depender de una aplicación o software preexistente. De esta forma, la herramienta queda como un elemento externo a herramientas de Ingeniería Ontológicas, facilitando la integración de la misma a sistemas enfocados en el diseño de ontologías, como por ejemplo *crowd*.

4. FORMACIÓN DE RECURSOS HUMANOS

En la Universidad Nacional del Comahue, Facultad de Informática, se otorgaron Becas CIN para estimular la vocación científica. Una de esas becas fue otorgada a uno de los autores de este trabajo, que está desarrollando su tesis de grado de la Licenciatura en Ciencias de la Computación en esta temática.

Por otra parte, otro de los autores de este trabajo es becario de investigación en la

²Listado disponible en

<http://ontologydesignpatterns.org/wiki/Category:ProposedContentOP>

categoría Graduado de Perfeccionamiento otorgada por la Universidad Nacional del Comahue y está desarrollando su tesis de Magíster en Ciencias de la Computación, posgrado que dicta la Facultad de Informática de la UNCo.

5. REFERENCIAS

- [1] F. Baader, I. Horrocks, C. Lutz, y U. Sattler. *An Introduction to Description Logic*. Cambridge University Press, 2017.
- [2] Germán Alejandro Braun, Christian Gimenez, Laura Andrea Cecchi, y Pablo R. Fillottrani. crowd: A Visual Tool for Involving Stakeholders into Ontology Engineering Tasks. *KI - Künstliche Intelligenz*, 2020.
- [3] M. Dudáš, S. Lohmann, V. Svátek, y D. Pavlov. Ontology visualization methods and tools: a survey of the state of the art. *The Knowledge Engineering Review*, 33, 2018.
- [4] Pablo Fillottrani, Enrico Franconi, y Sergio Tessaris. The ICOM 3.0 Intelligent Conceptual Modelling Tool and Methodology. *Semantic Web Journal*, 2012.
- [5] Nicola Guarino, Daniel Oberle, y Steffen Staab. What is an ontology? In Steffen Staab y Rudi Studer, editors, *Handbook on ontologies*, pages 1–17. Springer, 2009.
- [6] Terry Halpin y Tony Morgan. *Information modeling and relational databases*. Morgan Kaufmann, 2010.
- [7] Pascal Hitzler, Aldo Gangemi, y Krzysztof Janowicz. *Ontology engineering with ontology design patterns: foundations and applications*, volume 25. IOS Press, 2016.
- [8] Pascal Hitzler y Cogan Shimizu. Modular ontologies as a bridge between human conceptualization and data. In *International Conference on Conceptual Structures*, pages 3–6. Springer, 2018.
- [9] Ian Horrocks, Oliver Kutz, y Ulrike Sattler. The Even More Irresistible SROIQ. In *Proceedings of the Tenth International Conference on Principles of Knowledge Representation and Reasoning*, pages 57–67. AAAI Press, 2006.
- [10] C.M. Keet. *An Introduction to Ontology Engineering*. University of Cape Town, 2018.
- [11] Renars Liepinš, Mikus Grasmanis, y Uldis Bojars. OWLGrEd ontology visualizer. In *Proceedings of the 2014 International Conference on Developers*, volume 1268, pages 37–42. CEUR-WS. org, 2014.
- [12] Cogan Shimizu y Karl Hammar. CoModIDE–The Comprehensive Modular Ontology Engineering IDE. In *ISWC 2019 Satellite Tracks*, volume 2456, pages 249–252. CEUR-WS, 2019.
- [13] Cogan Shimizu, Quinn Hirt, y Pascal Hitzler. Modl: A modular ontology design library. *Preprint arXiv:1904.05405*, 2019.
- [14] Steffen Staab y Rudi Studer. *Handbook on ontologies*. Springer Science & Business Media, 2010.
- [15] York Sure, Steffen Staab, y Rudi Studer. Ontology engineering methodology. In Steffen Staab y Rudi Studer, editors, *Handbook on ontologies*, pages 135–152. Springer, 2009.
- [16] M. Vigo, S. Bail, C. Jay, y R. Stevens. Overcoming the Pitfalls of Ontology Authoring: Strategies yImplications for Tool Design. *International Journal of Human Computer Studies*, 2014.
- [17] Maleeha Arif Yasvi y Raghava Mutharaju. ODPReco-A Tool to Recommend Ontology Design Patterns. In *WOP@ ISWC*, pages 71–75, 2019.

Estrategias de Explotación de Información de Salud Pública en la Provincia del Chubut para su Uso en el Ordenamiento Territorial

Luciano Perdomo¹, Carlos Buckle^{1,2}, Denise Acosta⁵, Leo Ordinez^{1,2}, Ma. Florencia del Castillo^{3,4}

¹ Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco (FI-UNPSJB), Sede Trelew, Trelew, Argentina, +54 280-4428402

² Laboratorio de Investigación en Informática, Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco (LINVI-FI-UNPSJB), Puerto Madryn, Argentina, +54 280-4883585 Int. 117.

³ Instituto Patagónico de Ciencias Sociales y Humanas (IPCSH)- CCT-CENPAT -CONICET, Puerto Madryn, Argentina

⁴ Facultad de Humanidades y Ciencias Sociales, Universidad Nacional de la Patagonia San Juan Bosco

⁵ Hospital Zonal Puerto Madryn “Andrés Ísola”, Ministerio de Salud, Provincia del Chubut

lucianor.perdomo@gmail.com, leo.ordinez@gmail.com, cbuckle@unpata.edu.ar

Resumen

La pandemia COVID-19, ha impactado al mundo y su forma de vida. Por ser una enfermedad infecciosa de rápida propagación es necesario que se realice un seguimiento sobre su evolución y poder obtener evidencia para la toma de decisiones. Para resolver este problema, se presentó un proyecto de trabajo interdisciplinario, cuyo objetivo fue el desarrollo de un sistema de vigilancia, que permitiera hacer un seguimiento sobre los casos en la ciudad de Puerto Madryn, Chubut. El mismo, posee características tanto para realizar seguimiento en los eventos de salud, como para dar soporte a esa toma de decisiones en salud pública. El sistema denominado *RastreAr* articula diferentes dispositivos de salud pública, consolidando su información.

Palabras clave: Informática en Salud Pública, Epidemiología, Sistemas de vigilancia en salud.

Contexto

Este trabajo surge en el marco del proyecto de investigación *Análisis prospectivo inteligente del impacto social, económico y productivo*

del COVID-19 en la provincia de CHUBUT el cual fue presentado y aprobado en la convocatoria del Programa de Articulación y Fortalecimiento Federal de las capacidades en Ciencia y Tecnología Covid-19 del MINCYT. En dicho proyecto participan distintos centros e institutos de CONICET, la Universidad Nacional de la Patagonia San Juan Bosco con distintas dependencias, la Facultad Regional Chubut de la Universidad Tecnológica Nacional y otras instituciones como INTA y el Ministerio de Salud de Chubut.

El emprendimiento RastreAr tiene carácter interdisciplinario, enmarcado en un área que se denomina Informática en Salud Pública (PHI, Public Health Informatics) [1]. Este campo de investigación incluye a las ciencias de la computación y la información, la ingeniería de software, junto con la salud pública y la epidemiología, así como los estudios sociales y territoriales. En particular, el mismo se desarrolló en la ciudad de Puerto Madryn, más precisamente en articulación con el Hospital Zonal de Puerto Madryn “Andrés Ísola”.

Entre los principales objetivos de RastreAr, se encuentran la realización de un sistema de vigilancia, que permita obtener datos sobre la situación de casos de COVID-19 en la ciudad de Puerto Madryn (la cual supera los 100.000 habitantes); y la utilización de esos datos para

ayudar a las autoridades a tomar decisiones en cuanto a la salud pública. De esta manera, distintos dispositivos implementados de manera urgente, como son el programa nacional Detectar, la línea gratuita de atención telefónica de Chubut, el programa de Rastreadores del Hospital Zonal “Andrés Ísola”, los centros de testeos y los laboratorios privados, se consolidaron en un único punto central de información, para que todas las partes puedan acceder e intervenir, de acuerdo a sus capacidades.

1. Introducción

Debido a la pandemia COVID-19, en la ciudad de Puerto Madryn, provincia de Chubut, surge la necesidad de hacer un seguimiento sistematizado de los casos positivos de COVID, junto con sus contactos estrechos y los eventos asociados. Como subproducto, se considera realizar un análisis sobre los datos obtenidos para ayudar a la toma de decisiones. En este contexto, en el marco del proyecto “Análisis prospectivo inteligente del impacto social, económico y productivo del COVID-19 en la provincia de Chubut” aprobado para su financiamiento por el MINCYT en la convocatoria al Programa de articulación y fortalecimiento federal de las capacidades en ciencia y tecnología COVID-19, nace el proyecto “RastreAR” para poder crear un sistema y dar solución al problema planteado anteriormente. El mismo se encuentra en producción desde finales de abril del año pasado; en él se permite realizar altas de casos de COVID positivo, junto con su seguimiento y de sus contactos estrechos; además de la emisión de certificados y realización de encuestas. Se permite a los médicos, laboratorios y plan DetectAR ingresar pacientes como COVID positivo. Los rastreadores son los que realizan el seguimiento de los casos positivos y sus contactos estrechos. Los contactos estrechos se obtienen mediante encuestas automatizadas y por medio de las consultas de los rastreadores a los casos confirmados. Por otra parte, cabe resaltar que se ha realizado un análisis funcional, al comprender y dar

solución a los requisitos e incidentes que indican los usuarios sobre la utilización del sistema.

2. Motivación

La motivación que dió origen al proyecto fue la necesidad de realizar vigilancia epidemiológica sobre los eventos de COVID-19 en la ciudad de Puerto Madryn. Para poder abordarla, se emprendió un trabajo multidisciplinario entre epidemiólogos e informáticos. Otro aspecto que motivó este trabajo fue la construcción de conocimiento sobre PHI, con el objetivo de poder introducir la temática en el sistema sanitario provincial, para así poder ayudar a un mejor alcance del objetivo principal de la salud pública, que es mejorar la salud integral de la población.

3. Líneas de Investigación y Desarrollo

Según la OPS (Organización Panamericana de Salud), que forma parte de la OMS (Organización Mundial de la Salud) [2] los objetivos de la vigilancia en salud pública son a grandes rasgos, detectar, identificar, monitorear cambios, tendencias y patrones en la distribución de las enfermedades, en las prácticas y programas de salud de la población, así como planeación e investigación, control y prevención.

La presente investigación se apoya en un marco teórico interdisciplinario aún en consolidación, compuesto de diversos enfoques conceptuales, según sea el foco de la misma.

En primer lugar, las aplicaciones de la informática a la salud pública (Public Health Informatics, PHI, en inglés) surgen a mediados de los 90 [3] y principios de los 2000 [4]. Durante ese tiempo, el interés se centró en la construcción de una agenda que sintetice la utilidad de la PHI en términos del propio sistema sanitario, pero orientada a la solución tecnológica y no a la incorporación ubicua e imbricada de ambas disciplinas. En este sentido, la propuesta de 2018 de Wholey

et al. [5] aporta claridad en tanto define un perfil de Informático/a en Salud Pública desde el punto de vista del currículum académico. Los autores definen: “Los profesionales de la PHI son aquellos que trabajan en la práctica, la investigación o la academia y cuya función principal de trabajo es utilizar la informática para mejorar la salud de las poblaciones”. En línea con esta definición, surge de McFarlane *et al.* [6] la necesidad de “caracterizar a los especialistas en informática de salud pública (PHI) e identificar las necesidades de informática de la fuerza laboral de salud pública”. Este postulado implica también la demanda de construcción de mecanismos y técnicas de comunicación que permitan capturar los requerimientos de los/as agentes de salud, para ser interpretados por informáticos/as. Estos y otros tópicos emergentes, así como tendencias futuras, en el ámbito de la Informática en la Salud Pública han sido compendiadas de forma extensa por Magnuson y Dixon [1] y Miah *et al* [7].

Por otro lado, la inserción y aprovechamiento de la tecnología digital en el ámbito de la planificación urbana es un paradigma en construcción, que se ha sintetizado por Ratti y Claudel [8] bajo el nombre de Ciudades Inteligentes. Dentro de ese modelo, la Salud es uno de sus pilares [9], [10]; y gran parte de la investigación se centra en los servicios y equipamientos urbanos en relación a Salud Inteligente, en aplicaciones informáticas específicas, en la utilización de datos y en el diseño de infraestructuras acorde. En este sentido, el enfoque orientado a la solución tecnológica también se ve completado en cuanto a su impacto social [11], [12]. Más aún, [13], [14], [15] desde el LINVI se viene trabajando hace años en el acompasamiento de conceptos generales de las Ciudades Inteligentes, a los contextos locales patagónicos.

En términos del contexto de desarrollo del presente proyecto, la práctica de Salud Comunitaria genera acciones de índole territorial, en relación con prevención y asistencia, a partir de un acercamiento de los servicios de salud a las familias. Tanto en los sectores rurales como urbanos de Chubut, los

agentes de salud comunitaria ejercen su tarea cotidiana, realizando visitas regulares a las familias alcanzadas por una serie de programas sociales y sanitarios. Estas tareas de visita sistemática, se denominan rondas, y proveen al sistema de salud, una serie de datos referidos al “ambiente”; “individuo” y “hogar”. Actualmente esa información es recopilada en formato papel, con un bajo nivel y factibilidad de sistematización y cruce de datos.

4. Resultados esperados

Se espera como resultado general que el sistema pueda ser soporte de ayuda a los tomadores de decisiones en la salud pública, en la obtención de evidencia [16]; que permita realizar la vigilancia de manera eficaz; que la investigación realizada sea el punto de inicio y genere interés en los futuros profesionales, debido a que la región carece de especialistas de Informática en Salud Pública. No obstante, como se mencionó anteriormente, el presente proyecto se encuentra ya en producción y pleno funcionamiento. De este trabajo en progreso se destacan las siguientes características: desde diciembre de 2020 hasta marzo de 2021, momento en que se liberó la primera versión oficial a producción, se trabajó en el desarrollo de una aplicación que asistiera en el proceso de rastreo de Casos Confirmados y sus Contactos Estrechos. La misma integra todas las fuentes de información, junto con sus flujos de trabajo: Consultorios Clínicos, Programa Detectar, Laboratorios de Análisis Clínicos (LAC), Área Programática Norte, Rastreadores/as, Coordinadores/as de Rastreadores/as y Pacientes (confirmados y contactos estrechos). RastreAr permite el ingreso de los Casos Confirmados, sus Contactos Estrechos, desde diferentes fuentes (programa Detectar, LAC y consultorios clínicos); así como el registro de rastreos realizado individualmente por un equipo de más de 20 rastreadores/as. A la vez, genera certificados para los pacientes y correspondientes y colecta estadísticas para las autoridades sanitarias. El sistema RastreAr, al momento de la presentación de

esta nota, registra 74 usuarios, agrupados en 10 roles, 11.117 casos confirmados, 5.239 contactos estrechos y sobre los anteriores se han realizado 16.446 rastreos.

El sistema que comenzó con objetivo de realizar vigilancia epidemiológica; se ha ampliado para contener datos que puedan ser utilizados en la toma de decisiones en salud pública. Por ejemplo, se ha agregado que en las encuestas de pacientes, obligatoriamente deban indicar el barrio en el cual viven, para posterior monitoreo geográfico. Asimismo se ha incorporado un tablero de comandos que permite la visualización integral del estado de situación en una determinada fecha o período de tiempo (Casos confirmados, recuperados, contactos estrechos, fallecidos, detectados en laboratorios públicos y privados, detectados en Program Detectar, internados, UTI, etc).

Con todos estos datos, es posible realizar la vigilancia con uso en dos perspectivas: la primera es el seguimiento de eventos de salud; y la segunda, hacer vigilancia respecto a la salud pública. Los seguimientos con respecto a los eventos de salud, permiten monitorear, observar y detectar la ocurrencia de una enfermedad, su distribución, ver patrones y cambios, síntomas y comorbilidades de los pacientes, condición laboral, etc. Para la vigilancia en salud pública, la disponibilidad de estos datos en un eje de tiempo permite un mejor análisis y comprensión del estado de salud local de la ciudad de Puerto Madryn, sobre todo explotando la información geográfica, para lo cual existe un proyecto llamado *Mapyzer* [17], que permite integrar los totales por barrio y visualizarlos a través del tiempo.

5. Formación de recursos humanos

En este proyecto participan docentes investigadores del Departamento de Informática de la UNPSJB-Puerto Madryn, una médica epidemióloga en calidad de asesora y experta en el dominio y el desarrollo de las soluciones de software es abordado por un estudiante avanzado de la carrera de Licenciatura en Sistemas de Información de la

UNPSJB-Sede Trelew que realiza su tesina de grado en el marco de este proyecto.

Referencias

- [1] Magnuson, J. A., & Dixon, B. E. (Eds.). (2020). Public Health Informatics and Information Systems. Health Informatics. doi:10.1007/978-3-030-41215-9
- [2] Organización Panamericana de Salud. (2011). Módulos de principios de epidemiología para el control de enfermedades (MOPECE). All Type Assessoría Editorial Ltda . <https://www.paho.org/col/dmdocuments/MOP ECE4.pdf>
- [3] Friede, A., Blum, H. L., & McDonald, M. (1995). Public health informatics: how information-age technology can strengthen public health. *Annual review of public health*, 16, 239–252. <https://doi.org/10.1146/annurev.pu.16.050195.001323>
- [4] Brender, J., Nøhr, C., & McNair, P. (2000). Research needs and priorities in health informatics. *International Journal of Medical Informatics*, 58, 257-289. doi: 10.1016/s1386-5056(00)00092-7.
- [5] Wholey, D. R., LaVenture, M., Rajamani, S., Kreiger, R., Hedberg, C., & Kenyon, C. (2018). Developing workforce capacity in public health informatics: core competencies and curriculum design. *Frontiers in public health*, 6, 124. doi: 10.3389/fpubh.2018.00124
- [6] McFarlane, T. D., Dixon, B. E., Grannis, S. J., & Gibson, P. J. (2019). Public Health Informatics in Local and State Health Agencies: An Update From the Public Health Workforce Interests and Needs Survey. *Journal of public health management and practice : JPHMP*, 25 Suppl 2, Public Health Workforce Interests and Needs Survey 2017(2 Suppl), S67–S77. <https://doi.org/10.1097/PHH.0000000000000918>
- [7] Miah, Shah & Shen, Jun & Lamp, John & Kerr, Donald & Gammack, John.

- (2019). Emerging Insights of Health Informatics Research: A Literature Analysis for Outlining New Themes. *Australasian Journal of Information Systems*, 23. 10.3127/ajis.v23i0.2137.
- [8] Ratti, C., & Claudel, M. (2016). *The city of tomorrow: Sensors, networks, hackers, and the future of urban life*. Yale University Press.
- [9] Pacheco Rocha, Dias, Santinha, Rodrigues, Queirós, & Rodrigues. (2019). Smart Cities and Healthcare: A Systematic Review. *Technologies*, 7(3), 58. doi:10.3390/technologies7030058
- [10] Buttazzoni, A., Veenhof, M., & Minaker, L. (2020). Smart City and High-Tech Urban Interventions Targeting Human Health: An Equity-Focused Systematic Review. *International journal of environmental research and public health*, 17(7), 2325. <https://doi.org/10.3390/ijerph17072325>
- [11] González, V. R., & Revolware, A. J. (2015). Smart, Health, Smart-City y Poder: la vida del nuevo ciudadano. In *El mejoramiento humano: avances, investigaciones y reflexiones éticas y políticas* (pp. 185-193).
- [12] Tan, S., & Taeihagh, A. (2020). Smart City Governance in Developing Countries: A Systematic Literature Review. *Sustainability*, 12(3), 899. doi:10.3390/su12030899
- [13] Sanchez, A., Ordinez, L., Firmenich, S., Barry, D., & Santos, R. (2017). An Expert-driven Ontology-based Approach to the Collaborative Acquisition of Information. *Journal of Computer Science and Technology*, 17(02), e17. <https://doi.org/10.24215/16666038.17.e17>
- [14] Sánchez, A., & Ordinez, L. (2020). Towards an ontology for designing cycling routes in the city of Puerto Madryn. In *2020 Seventh International Conference on eDemocracy eGovernment (ICEDEG)* (pp. 143-150).
- [15] Leo Ordinez, Carlos Buckle, Sergio Andrés Kaminker, Diego Firmenich, Damián Barry, & Ariel Aguirre (2021). Assessing cycling social feasibility in a medium-size Patagonian city. *Transportation Research Part D: Transport and Environment*, 92, 102720.
- [16] Brownson RC, Fielding JE, Maylahn CM. (2009). Evidence-based public health: a fundamental concept for public health practice. *Annu Rev Public Health*. 2009;30:175-201. doi: 10.1146/annurev.publhealth.031308.100134.
- [17] Nuñez, G. M., Jaureguibehe, M., Buckle, C., Ordinez, L., & Barry, D. (2021). Mapyzer: una herramienta de carga y visualización de datos espacio-temporales. In *XXVII Congreso Argentino de Ciencias de la Computación (CACIC)(Modalidad virtual, 4 al 8 de octubre de 2021)*.

PDP – Procesamiento Distribuido y en Paralelo

Inteligencia artificial para la multi-clasificación de fauna en fotografías automáticas utilizadas en investigación científica

Gonzalez Federico¹, Viera Leonel¹, Soler Rosina², Chiarvetto Peralta Lucila¹, Gel Matías¹, Bustamante Gimena², Montaldo Abril¹, Rigoni Brian¹, Perez Ignacio¹

¹ Instituto de Desarrollo Económico e Innovación
Universidad Nacional de Tierra del Fuego, Antártida e Islas del Atlántico Sur
IDEI - UNTDF
Ushuaia, Tierra del Fuego

² Centro Austral de Investigaciones Científicas
CADIC - CONICET
Ushuaia, Tierra del Fuego

{fgonzalez, lviera, mgel, lchiarvetto}@untdf.edu.ar
{rosina.soler, gimenabustamante}@conicet.gov.ar
{abril.4545, brian.rigoni1, ignacioperez583}@gmail.com

RESUMEN

El manejo de ambientes naturales, ya sea para conservación o producción, requiere de una profunda comprensión de la vida silvestre. El número, la ubicación y el comportamiento de los animales salvajes es uno de los principales objetos de estudio en ecología y vida silvestre. El uso de cámaras trampa ofrece la oportunidad de recopilar rápidamente grandes cantidades de fotografías que -sin la presencia humana- registran a la fauna en su hábitat natural, evitando factores que alteren su comportamiento.

En Tierra del Fuego, Argentina, se desarrollan investigaciones sobre el uso del bosque por parte de distintos herbívoros (guanacos, vacas, ovejas) para optimizar el manejo y proteger dichos ecosistemas naturales.

Si bien las cámaras trampa permiten la obtención de millones de imágenes, la interpretación de tales fotografías representa un problema de escala para el procesamiento manual. Así, gran parte del valioso conocimiento en estos enormes repositorios de datos sigue sin ser aprovechado.

Las Redes Neuronales y el Deep Learning son áreas de estudio dentro la Inteligencia Artificial, durante la última década estas dos disciplinas han hecho cuantiosos aportes en el ámbito del reconocimiento de imágenes de gran relevancia a nivel mundial.

Los estudios ecológicos y de conservación de la vida silvestre, pueden combinarse con estas nuevas tecnologías para extraer información importante a partir de las fotografías obtenidas por cámaras trampa, con el objeto de aportar a la comprensión de distintos procesos naturales y mejorar el manejo de las áreas silvestres implicadas.

Nuestro proyecto busca desarrollar modelos de redes neuronales para clasificar especies de animales en fotografías obtenidas mediante cámaras trampa, para resolver problemas de gran volumen en investigación científica.

Palabras clave: deep learning ; computer vision ; trap camera ; vida silvestre

CONTEXTO

Este proyecto forma parte de la línea de investigación "Innovación en sistemas de software" del área 8 "Desarrollo Informático", del Instituto de Desarrollo Económico e Innovación de la Universidad Nacional de Tierra del Fuego (IDEI-UNTDF).

Nuestro grupo de investigación viene trabajando temas afines desde 2017 y en 2020 comenzamos con la interpretación de cámaras trampa en el proyecto interno del IDEI "Inteligencia Artificial para la identificación de fauna en fotografías automáticas utilizadas en investigación científica". Recientemente,

hemos postulado a un proyecto PID-UNTDF, que esperamos sea aprobado próximamente y su ejecución se dará a lo largo de 2022 y hasta mediados de 2023.

INTRODUCCIÓN

El manejo de ambientes naturales, ya sea para conservación o producción, requiere el conocimiento del comportamiento de la vida silvestre. El número, la ubicación y el comportamiento de los animales silvestres es uno de los principales objetos de estudio en ecología y manejo de la vida silvestre (Silveira et al. 2003, Fegraus et al. 2011, Palmer y Packer 2018). El uso de cámaras con sensores de movimiento en hábitats naturales -llamadas cámaras trampa- ha transformado la investigación en ecología y conservación de la vida silvestre en las últimas dos décadas (O'Connell et al. 2010). Estas se han convertido en una herramienta esencial para los ecologistas, permitiéndoles estudiar el tamaño y la distribución de las poblaciones (Silveira et al. 2003) y evaluar el uso del hábitat. Si bien permiten tomar millones de imágenes (Fegraus et al. 2011), la clasificación y extracción de información (datos) es tradicionalmente realizada por humanos (es decir, expertos o una comunidad de voluntarios), es lenta y costosa debido a su procesamiento manual. Así, gran parte del valioso conocimiento en estos grandes repositorios de datos sigue sin ser aprovechado.

Actualmente, estas cámaras tienen sensores de movimiento que se activan con la presencia de un animal, registrando una imagen digital por cada situación detectada. Dichos sensores pueden configurarse para ser más o menos sensibles. Técnicamente, el sensor de movimiento reacciona a los cambios de luces, una configuración sensible produce una mayor cantidad de falsos positivos, una configuración menos sensible es propensa a no fotografiar todas las situaciones necesarias.

Para evitar el sub-muestreo de individuos, es común configurar la cámara en un rango de mayor sensibilidad; esto lleva a obtener una inmensa cantidad de registros innecesarios, ya que cualquier movimiento será interpretado

como "necesario de fotografiar". Así, por ejemplo, dentro de un bosque los movimientos de ramas u otra vegetación activan el sensor y la cámara acciona el diafragma, pero probablemente ningún animal haya estado presente. Lo mismo ocurre cuando una nube tapa el sol repentinamente, cuando nieva, etc. Otro de los problemas típicos es la presencia de animales que no son de interés para el estudio (ej, aves), que inevitablemente se cruzan delante del lente y accionan la cámara. Finalmente, muchas de estas cámaras tienen capacidad de hacer fotografías nocturnas gracias a tecnología de infrarrojos, que si bien producen imágenes en blanco y negro, son de gran utilidad, ya que muchos animales tienen hábitos nocturnos.

Las cámaras trampa pueden pasar meses en el campo trabajando automáticamente, cuentan con baterías de larga duración y con memorias digitales de gran capacidad. Dependiendo del ambiente donde se instale el equipo, de las condiciones climáticas, el tipo de animales esperados y la sensibilidad configurada, es común que una de cada mil fotografías tomadas sea de interés para el estudio en cuestión. Si a esto se le suma la cantidad de tiempo que la cámara pase activa en el campo (semanas, meses) y la cantidad de cámaras trabajando en paralelo dentro de una misma investigación, el procesamiento de las imágenes y la tarea de encontrar manualmente las fotografías útiles para el estudio, puede convertirse en un trabajo abrumador o incluso imposible.

La Inteligencia Artificial (IA) ha tenido un importante crecimiento y popularidad durante la última década, y particularmente durante los últimos cinco años, gracias a un innovador enfoque que dio impulso al Deep Learning y las Redes Neuronales (Goodfellow et al. 2016, He et al. 2016), junto con una importante capacidad de procesamiento gracias a la invención de las GPU (Graphics Processing Unit).

Los modelos basados en redes neuronales con aprendizaje profundo son capaces de identificar patrones en imágenes, obtener conclusiones basadas en un denominador común y generar un modelo que representa a

cierto grupo de imágenes con un patrón similar (Simonyan y Zisserman 2014). Luego, dada una imagen cualquiera, dicho modelo puede utilizarse para reconocer si esa imagen coincide en su representación con su patrón definido.

Los estudios ecológicos y de conservación de la vida silvestre, pueden combinarse con la IA para extraer información importante a partir de las fotografías obtenidas por cámaras trampa, con el objeto de aportar a la comprensión de distintos procesos naturales y mejorar el manejo de las áreas silvestres implicadas.

Actualmente, se encuentra en desarrollo el proyecto “Evaluación del impacto diferencial de herbívoros nativos y domésticos en bosques de *Nothofagus antarctica* (ñire) con producción ganadera en Tierra del Fuego” a cargo de la Dra. Rosina Soler, investigadora del Laboratorio de Recursos Agroforestales del CADIC-CONICET en Ushuaia. El proyecto se desarrolla en bosques nativos del centro de la provincia, en cercanías del Municipio de Tolhuin. El estudio analiza la interacción planta-herbívoro en los bosques de ñire, a partir de la respuesta de la regeneración arbórea (densidad, crecimiento, supervivencia) a la exclusión del ganado y de los herbívoros nativos. El proyecto ha recolectado más de 150.000 fotos en el período 2015-2017 utilizando un set de 24 cámaras trampa.

Esta cantidad de información representa un alto costo de análisis manual de imágenes, ralentiza el trabajo y dificulta la obtención de resultados. Durante 2020 y parte de 2021 se desarrolló también el proyecto “Inteligencia Artificial para la identificación de fauna en fotografías automáticas utilizadas en investigación científica”, bajo la dirección de Federico Gonzalez y con prácticamente el mismo grupo de trabajo de este proyecto. Con la ayuda del Deep Learning y la IA, el proyecto permitió separar fotos “con animales” y “sin animales” de la mencionada base de datos.

La nueva etapa de estudio que abordaremos próximamente, pretende conocer más detalles sobre las especies encontradas en cada una de las 150.000 imágenes. Ya no sólo limitarse a clasificar presencia-absencia de animal, sino identificar qué tipo de especies hay en cada

imagen y cuántos individuos de cada una.

Más allá de lo positivo que este proyecto podría tener en el plano local, los resultados obtenidos de esta colaboración que se plantea entre el CADIC y la UNTDF tendrían la capacidad de ser transferibles globalmente a otros estudios similares en los que se utilicen este tipo de cámaras y donde se requiera clasificar imágenes para identificar fauna silvestre.

Finalmente, también cabe destacar que en la carrera de Licenciatura en Sistemas de la UNTDF no existe ninguna materia que trate temas de IA en profundidad, por lo tanto los alumnos involucrados en este proyecto de investigación tendrán la posibilidad de obtener conocimiento de gran valor para su vida profesional.

LÍNEAS DE INVESTIGACIÓN

Nuestro trabajo se enmarca en la Inteligencia Artificial como una gran línea de investigación, aunque nuestro foco de atención se encuentra específicamente en el Deep Learning (aprendizaje profundo) y en Computer Vision (visión por computador) como ejes específicos.

RESULTADOS Y OBJETIVOS

Nuestro proyecto busca desarrollar modelos de redes neuronales basados en Deep Learning para clasificar animales en fotografías obtenidas mediante cámaras trampa, haciendo uso de la IA para resolver problemas de gran volumen en investigación científica.

En 2021 hemos podido clasificar la base de datos con 150.000 imágenes, identificando en cada una la presencia-absencia de animales con una precisión del 85% aprox. y con una variación según los ambientes naturales, la época del año y la luz disponible (día y noche) de $\pm 10\%$.

Nuestro próximo objetivo pretende especializar el modelo anterior, para clasificar las imágenes según especie animal, con un nivel de precisión similar al que una persona entrenada podría alcanzar mediante separación manual.

Luego esperamos desarrollar un modelo de red

neuronal capaz de generalizarse a otros ambientes naturales y otras especies animales. Por lo cual, los resultados obtenidos tendrían la capacidad de ser transferibles globalmente a otros estudios similares.

Desde un punto de vista académico se espera aportar al conocimiento de los docentes-investigadores que participan directamente del proyecto, quienes actualmente se encuentran realizando estudios de Doctorado y Maestría; pero también para los alumnos participantes quienes manifiestan interés en aplicarlos en sus propuestas de tesis de licenciatura.

Finalmente, se espera volcar localmente lo aprendido en un taller de presentación de resultados, que estará orientado a profesores de la UNTDF, investigadores del CADIC-CONICET y otros profesionales interesados.

FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está compuesto por personas provenientes de diferentes disciplinas y antecedentes. En el campo de la educación, tres profesores adjuntos con cargo docente-investigador, un asistente y tres estudiantes: todos ellos vinculados a la carrera de Lic. en Informática en la UNTDF.

A su vez, por parte de CADIC-CONICET, participan la Dra. Rosina Soler y la Ing. Gimena Bustamante. La Dra. Rosina Soler es bióloga, investigadora adjunta de CONICET y dirige el proyecto “Evaluación del impacto diferencial de herbívoros nativos y domésticos en bosques de *Nothofagus antarctica* (ñire) con producción ganadera en Tierra del Fuego”, desde donde se genera la base de datos que utilizamos en este trabajo. La Ing. Gimena Bustamante está finalizando su Doctorado en Ciencias Agrarias y Forestales en la UNLP.

Mientras que la arista informática cuenta con perfiles de diferentes especialidades. El Lic. Federico González quien dirige el proyecto, tiene un Máster en Ciudades Inteligentes por la Univ. de Girona (UdG), España y está desarrollando su Doctorado en Inteligencia Artificial en la Univ. de Barcelona (UB).

Mientras que el Lic. Leonel Viera -co-director del proyecto- tiene la expectativa de finalizar

su Máster en Inteligencia de Datos y BigData en la Univ. Nacional de La Plata (UNLP) dentro del marco de nuestro proyecto.

El Lic. Matías Gel está también desarrollando su Máster en Inteligencia de Datos y BigData en la UNLP.

La Lic. Lucila Chiarvetto Peralta está finalizando su Doctorado en Ciencias de la Computación en la Univ. Nacional del Sur (UNS).

Por último, los alumnos Abril Montaldo, Ignacio Perez y Brian Rigoni son Analistas Universitarios en Sistemas y están cursando su último año de Lic. en Informática, de quienes se espera que puedan comenzar su tesina de grado en temas afines al proyecto.

REFERENCIAS

Deng J, Dong W, et al. (2009) Imagenet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, pp. 248-255.

Fegraus EH, Lin K, et al. (2011) Data acquisition and management software for camera trap data: A case study from the team network. *Ecol Inform* 6(6): 345-353.

Goodfellow I, Bengio Y, Courville A (2016) *Deep Learning*. MIT Press, Cambridge, MA, USA.

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp.

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. 2012 Advances in Neural Information Processing Systems (NIPS).

LeCun YA, Bottou L, Orr GB, Müller KR (2012) Efficient backprop in Neural networks: Tricks of the trade. In: Montavon G, Orr GB, Müller KR. (eds) *Lecture Notes in Computer Science*, vol 7700. Springer, Berlin.

Stuart J. Russell, Peter Norvig (2010) *Artificial Intelligence: A Modern Approach*, Third Edition, Prentice Hall ISBN 9780136042594.

O'Connell AF, Nichols JD, Karanth KU (2010) *Camera Traps in Animal Ecology: Methods and Analyses*. Springer, Tokyo, Japan.

Palmer MS, Packer C (2018) Giraffe bed and breakfast: Camera traps reveal Tanzanian yellow-billed oxpeckers roosting on their large mammalian hosts. *Afr J Ecol* 56(4): 882-884.

Silveira L, Jacomo AT, Diniz-Filho JAF (2003) Camera trap, line transect census and track surveys: A comparative evaluation. *Biol Conserv* 114(3): 351-355.

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556

Szegedy C, Liu W, et al. (2015) Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp.

Wiesler S, Ney H (2011) A convergence analysis of log-linear training. 2011 Advances in Neural Information Processing Systems (NIPS).

Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? 2014 Advances in Neural Information Processing Systems (NIPS).

Desarrollo de un simulador para la evaluación de algoritmos clásicos y nuevos para la gestión de recursos compartidos en sistemas distribuidos contemplando exclusión mutua

David L. La Red Martínez, Stella Maris Gerzel, Federico Agostini, Manuel Alejandro Ricardone, Cynthia Evelin Bruic, Hugo R. Haurech, Carlos Leandro M. Latyn

Universidad Nacional del Chaco Austral
Sáenz Peña, (3700), Argentina

laredmartinezdavid@uncaus.edu.ar, stellagerzel@uncaus.edu.ar, agostinifede@hotmail.com,
manuelricardone@uncaus.edu.ar, Cynthia.bruic@gmail.com, hugohaurech@gmail.com,
leandrolatyn@uncaus.edu.ar

Resumen

En los sistemas de procesamiento distribuido es necesario que los procesos que actúan en grupos deban tomar decisiones basados en acuerdos respecto del acceso a recursos; las decisiones pueden estar relacionadas con la realización de determinada actividad que requiera o no la sincronización de los procesos, es decir, que los procesos del grupo estén activos en los mismos lapsos en sus respectivos procesadores, requiriendo el uso de recursos compartidos en la modalidad de exclusión mutua mediante consensos estrictos o no. Así surge el siguiente interrogante: ¿Cuáles son los modelos de decisión y los operadores de agregación que habrá que generar incorporando la perspectiva cognitiva a los modelos clásicos para la toma de decisiones en la gestión de grupos de procesos, que trasciendan el enfoque tradicional de las ciencias de la computación, teniendo en cuenta la autorregulación? ¿Cómo se implementarán los algoritmos de los distintos modelos de decisión? ¿Cómo validar los nuevos algoritmos propuestos comparándolos entre sí y con los algoritmos tradicionales? Para ello habrá que desarrollar un simulador que implemente los algoritmos tradicionales y los nuevos propuestos y permita observar su comportamiento y resultados ante diferentes tipos de cargas de trabajo. Estas actividades se

desarrollan en el marco del PI N° 126, aprobado por Res. N° 011/20 CS de la UNCAus.

Palabras Clave: operadores de agregación, sistemas operativos, planificación de procesadores, simulador de planificación de procesos

Contexto

Debido a la proliferación de las redes de comunicaciones de datos es cada vez mayor la utilización de sistemas distribuidos en los cuales los distintos procesos compiten por el uso de recursos, muchos de ellos compartidos y accedidos en la modalidad de exclusión mutua para asegurar integridad de las estructuras de datos o debido a la naturaleza del recurso, lo cual impacta en el rendimiento al limitar el paralelismo. Esta problemática ha sido estudiada ampliamente y sus resultados constituyen los llamados modelos clásicos, entendiendo por tales al algoritmo centralizado, al algoritmo distribuido de Lamport, Ricart y Agrawala, al algoritmo de anillo de fichas, entre otros.

Las soluciones mencionadas precedentemente no tienen en cuenta una visión global del sistema distribuido ni del estado de los distintos nodos de procesamiento y

almacenamiento en particular, considerando las distintas variables que configuran su carga computacional.

Surge así la necesidad de generar nuevos “modelos de decisión” y “operadores de agregación” que incorporen la perspectiva cognitiva para la toma de decisiones en la gestión de grupos de procesos, teniendo en cuenta la autorregulación. Se debe validar los nuevos modelos generados comparándolos entre sí y con los tradicionales. Para ello se propone desarrollar un simulador que implemente los algoritmos tradicionales y los nuevos propuestos y permita observar su comportamiento y resultados ante diferentes tipos de cargas de trabajo.

La realización del simulador mencionado por sí solo constituiría un aporte importante para la enseñanza de los sistemas operativos, especialmente los sistemas distribuidos, en tanto que los nuevos modelos de decisión y los algoritmos y operadores de agregación que los integren podrían constituir un significativo aporte al mejoramiento del rendimiento global de los sistemas distribuidos en los cuales se comparten recursos, muchos de ellos en la modalidad de exclusión mutua distribuida.

Introducción

En los sistemas informáticos, muchos de ellos distribuidos, en los cuales existen múltiples procesos que cooperan para el logro de una determinada función, es necesario disponer de modelos de decisión que permitan a los procesos intervinientes en los distintos grupos de procesos, tomar decisiones en las que son necesarios diferentes niveles de acuerdo, especialmente cuando se trata del acceso a recursos computacionales compartidos y el sistema debe autorregular la forma de dicha compartición.

Es especialmente significativo el caso del acceso a las llamadas regiones críticas de memoria por parte de distintos procesos, que pueden estar operando en equipos distribuidos,

donde el acceso a las regiones críticas debe hacerse en la modalidad de acceso exclusivo y con el consentimiento de los demás procesos del grupo.

Ejemplos de lo mencionado se encuentran en [1] y [2], donde se describen los principales algoritmos de sincronización en sistemas distribuidos, en [3], donde se presenta una solución eficiente y tolerante a fallas para el problema de la exclusión mutua distribuida, en [4], [5] y en [6], donde se presentan unos algoritmos para gestionar la exclusión mutua en redes de computadoras, en [7], donde se describen los principales algoritmos de sincronización en sistemas distribuidos, en [8], donde se detallan los principales algoritmos para la gestión distribuida de procesos, los estados globales distribuidos y la exclusión mutua distribuida. Estos temas y otros relacionados también han sido tratados en [9], [10], etc.

Los modelos de decisión disponibles en la actualidad y generalmente aplicables en los sistemas distribuidos se basan en algoritmos de intercambio de permisos que intentan lograr un acuerdo de todos los procesos intervinientes para realizar determinadas acciones, como el acceso a un área de memoria compartida a la que se debe acceder en la modalidad de exclusión mutua.

En los sistemas computacionales de procesamiento distribuido es frecuentemente necesario que los procesos que actúan en grupos deban tomar decisiones basados en el acuerdo; dichos procesos podrán operar en un mismo equipo informático o en varios equipos distribuidos interconectados; las decisiones para las cuales deben alcanzar algún nivel de acuerdo pueden estar relacionadas con la realización de determinada actividad que no requiera el uso de recursos compartidos en la modalidad de exclusión mutua, o con la realización de determinada actividad que sí requiera el uso de recursos compartidos en la modalidad de exclusión mutua, para lo cual

generalmente las exigencias de niveles de acuerdo son mayores que para el caso anterior, pudiendo darse además que los procesos integren grupos que requieran (o no) sincronización (estar activos en sus respectivos procesadores en un mismo lapso de tiempo).

Los principios de la cibernética de segundo orden, los sistemas complejos y la autorregulación, posibilitan desarrollar modelos de decisión desde la óptica cognitiva para la toma de decisiones en grupos de procesos, que trasciendan el enfoque tradicional de las ciencias de la computación considerando la posibilidad de imputación de datos faltantes y la fuzzyficación de ciertas variables, utilizando la familia de operadores OWA, generando operadores específicos para cada uno de los siguientes tipos de situaciones:

- Que los procesos accedan a recursos compartidos en la modalidad de exclusión mutua sin constituir grupos de procesos que requieran sincronización (estar activos en sus respectivos procesadores en un mismo lapso) y con exigencias estrictas de consenso para lograr el acceso.
- Que los procesos accedan a recursos compartidos en la modalidad de exclusión mutua sin constituir grupos de procesos que requieran sincronización (estar activos en sus respectivos procesadores en un mismo lapso) y sin exigencias estrictas de consenso para lograr el acceso.
- Que los procesos accedan a recursos compartidos en la modalidad de exclusión mutua constituyendo grupos de procesos que requieren sincronización (estar activos en sus respectivos procesadores en un mismo lapso) y con exigencias estrictas de consenso para lograr el acceso.
- Que los procesos accedan a recursos compartidos en la modalidad de exclusión mutua constituyendo grupos de procesos que requieren sincronización (estar activos en sus respectivos procesadores en un mismo lapso) y sin exigencias estrictas de consenso para lograr el acceso.

Líneas de Investigación y Desarrollo

Se considera especialmente importante estudiar la aplicación de modelos de decisión para la toma de decisiones en grupo que se desprendan de conceptos cognitivos de la cibernética en general y de la cibernética de segundo orden en particular, en el contexto de sistemas complejos autorregulados.

Se estima de gran importancia estudiar los mecanismos de autorregulación de los sistemas, especialmente de los sistemas complejos, en el contexto de la cibernética de segundo orden, a los efectos de su posterior incorporación en el desarrollo de modelos de decisión aplicables a procesos distribuidos que deben tomar decisiones en grupo respecto del uso de recursos compartidos, con requisitos de sincronización. Se considera en tal sentido que dichos grupos de procesos mejorarían su desempeño mediante los modelos de decisión que se tiene previsto desarrollar incorporando mecanismos de autorregulación y conceptos de la cibernética de segundo orden en el proceso de toma de decisiones. Se pretende generar nuevos modelos de toma de decisiones en grupos de procesos distribuidos, contemplando además la aplicación de métodos de imputación de datos para aquellos casos de datos faltantes, por ejemplo, como consecuencia de problemas en las comunicaciones entre los procesos, y fuzzyficación de variables para dar soporte a situaciones donde no es posible o conveniente expresar valores exactos.

Habrà que desarrollar, por lo tanto, los modelos de decisión para la toma de decisiones en grupos de procesos, para los siguientes tipos de situaciones: a) que los procesos accedan a recursos compartidos en la modalidad de exclusión mutua sin constituir grupos de procesos que requieran sincronización (estar activos en sus respectivos procesadores en un mismo lapso de tiempo) y con exigencias estrictas de consenso para lograr el acceso; b) que los procesos accedan a recursos

compartidos en la modalidad de exclusión mutua sin constituir grupos de procesos que requieran sincronización (estar activos en sus respectivos procesadores en un mismo lapso de tiempo) y sin exigencias estrictas de consenso para lograr el acceso; c) que los procesos accedan a recursos compartidos en la modalidad de exclusión mutua constituyendo grupos de procesos que requieren sincronización (estar activos en sus respectivos procesadores en un mismo lapso de tiempo) y con exigencias estrictas de consenso para lograr el acceso; d) que los procesos accedan a recursos compartidos en la modalidad de exclusión mutua constituyendo grupos de procesos que requieren sincronización (estar activos en sus respectivos procesadores en un mismo lapso de tiempo) y sin exigencias estrictas de consenso para lograr el acceso.

Atento a lo indicado precedentemente, y además de las diferencias cualitativas que presentarán los diferentes algoritmos, es necesario evaluar su comportamiento mediante un simulador de cargas de trabajo, el que tendrá que ser desarrollado para simular el comportamiento de los diferentes algoritmos para diferentes cargas de trabajo.

Como consecuencia del análisis, podría ser necesario modificar los modelos de decisión propuestos, lo cual iniciaría un nuevo ciclo de validaciones.

Resultado Esperados

Definir teóricamente los modelos de decisión mencionados, proceder a la validación de estos comparando sus prestaciones con las de los modelos de las ciencias de la computación habitualmente utilizados en los sistemas operativos, para lo cual se desarrollará un simulador específico.

En el marco del PI se ha realizado una publicación relacionada con un nuevo método para la selección de recursos en la nube [11] y

se ha desarrollado una primera versión del simulador, que se describe en [12].

Formación de Recursos Humanos

El equipo de trabajo está compuesto por un Doctor, dos Magisters, una Licenciada en Sistemas de Información, una Ingeniera de Sistemas y un estudiante avanzado en condición de adscripto al proyecto.

Referencias

- [1] Tanenbaum, A. S. *Sistemas Operativos Distribuidos*. Prentice-Hall Hispanoamericana S.A. México. 1996.
- [2] Tanenbaum, A. S. *Sistemas Operativos Modernos*. 3ra. Edición. Pearson Educación S. A. México. 2009.
- [3] Agrawal, D. y El Abbadi, A. An Efficient and Fault-Tolerant Solution of Distributed Mutual Exclusion. *ACM Trans. on Computer Systems*. Vol. 9. Pp. 1-20. USA. 1991.
- [4] Ricart, G. y Agrawala, A.K. An Optimal Algorithm for Mutual Exclusion in Computer Networks. *Commun. of the ACM*. Vol. 24. Pp. 9-17. 1981.
- [5] Cao, G. and Singhal, M. A Delay-Optimal Quorum-Based Mutual Exclusion Algorithm for Distributed Systems. *IEEE Transactions on Parallel And Distributed Systems*. Vol. 12, no. 12. Pp. 1256-1268. USA. 2001.
- [6] Lodha, S. and Kshemkalyani, A. A Fair Distributed Mutual Exclusion Algorithm. *IEEE Trans. Parallel and Distributed Systems*. Vol. 11. N° 6. Pp. 537-549. USA. 2000.
- [7] La Red Martínez, D.L. *Sistemas Operativos*. EUDENE. Argentina. 2004
- [8] Stallings, W. *Sistemas Operativos*. 5ta. Edición. Pearson Educación S.A. España. 2005.
- [9] Joshi, R., Holzmann, G. J. A Mini-Challenge: Build a Verifiable Filesystem, *Formal Aspects of Computing*, Vol. 19. 2007.

[10] Alagarsamy, K. Some Myths About Famous Mutual Exclusion Algorithms. ACM SIGACT News 34 (3): 94–103. 2003.

[11] Haurech, H.R.; La Red Martínez, D. L. The Analytic Hierarchy Process as a Method for the Selection of Resources in the Cloud., en F. V. Cipolla-Ficarra (Ed.), Handbook of Research on Software Quality Innovation in Interactive Systems. Hershey; ISSN N° 2327-3453; PA, USA: IGI Global Engineering Science Reference; 2021.

[12] D. L. la Red Martínez, F. Agostini, J. C. Acosta, S. Gerzel, L. Latyn; “Simulador para la evaluación de algoritmos para la gestión de recursos compartidos en sistemas distribuidos”; Revista de Investigación en Tecnologías de la Información (RITI); ISSN: 2387-0893; Barcelona, España (Aceptado por el editor y en etapa de evaluación externa, aún no publicado).

DISEÑO DE METAHEURÍSTICAS PARALELAS CON EL PARADIGMA *NOVELTY SEARCH* PARA LA REDUCCIÓN DE INCERTIDUMBRE EN LA PREDICCIÓN DE FENÓMENOS DE PROPAGACIÓN

Jan Strappa^{ab}, Paola Caymes Scutari^{ab} y Germán Bianchini^a

^aLaboratorio de Investigación en Cómputo Paralelo/Distribuido (LICPaD),
Dpto. de Ingeniería en Sistemas de Información - UTN-FRM - Mendoza - Argentina.

^bConsejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

jstrappa@frm.utn.edu.ar, pcaymesscutari@frm.utn.edu.ar, gbianchini@frm.utn.edu.ar

RESUMEN

Los incendios forestales son un fenómeno ambiental multicausal de gran prevalencia. El impacto de este fenómeno incluye pérdidas humanas, daños ambientales y económicos. Para mitigar estos daños, existen sistemas de simulación computacionales que predicen el comportamiento del fuego en base a un conjunto de parámetros de entrada o escenario (velocidad, dirección del viento; temperatura; etc.). Sin embargo, los resultados de una simulación suelen tener un alto grado de error por la incertidumbre en los valores de algunas variables, por no ser conocidos o porque su medición puede ser imprecisa o errónea. Por este motivo se han desarrollado métodos que combinan resultados de un conjunto de simulaciones sobre distintos escenarios, para detectar tendencias y así reducir dicha incertidumbre. Dos propuestas recientes, ESSIM-EA y ESSIM-DE, utilizan algoritmos evolutivos paralelos para orientar el espacio de escenarios a considerar, logrando mejoras en la calidad predictiva. Estos enfoques están guiados por una función objetivo que recompensa el avance hacia una solución. En problemas complejos, dicha función objetivo no siempre es un indicador directo de la calidad de las soluciones. En trabajos previos se han encontrado limitaciones como convergencia prematura, y se han requerido acciones de calibración y sintonización para incorporar soluciones más diversas al proceso de

predicción. Para superar estas limitaciones, en este trabajo proponemos aplicar el paradigma *Novelty Search* (búsqueda basada en novedad), que reemplaza la función objetivo por una medida de la novedad de las soluciones encontradas, para generar continuamente soluciones con comportamientos diferentes entre sí. Este enfoque logra evitar óptimos locales y permitiría encontrar soluciones útiles que serían difíciles de hallar por otros algoritmos. Al igual que los métodos existentes, esta propuesta también puede aplicarse a otros modelos de propagación (inundaciones, avalanchas o corrimientos de suelo).

Palabras Clave: Predicción de Incendios Forestales, Metaheurísticas, Búsqueda Basada en Novedad, Reducción de Incertidumbre.

CONTEXTO

El proyecto aquí descrito se está llevando a cabo en el marco de una línea de investigación que se ha desarrollado a lo largo de los últimos años en el LICPaD (Laboratorio de Investigación en Cómputo Paralelo/Distribuido), en la UTN-FRM. Esta línea involucra el desarrollo de un sistema de predicción del comportamiento de fenómenos de propagación, aplicado al caso particular del avance de la línea de fuego en incendios

forestales. Este tipo de sistema utiliza un modelo de simulación que toma como entrada un conjunto de parámetros, que describen características que afectan a la propagación. Los métodos desarrollados por el grupo tienen como objetivo la reducción de incertidumbre en estas simulaciones, dado que las limitaciones para proveer parámetros de entrada correctos para el modelo conducen a errores en la precisión de la predicción. Recientemente se han desarrollado sistemas que utilizan métodos evolutivos para obtener predicciones basadas en un conjunto de simulaciones. En este proyecto, se propone el uso de una metaheurística dentro de un paradigma de búsqueda alternativo, llamado *Novelty Search* o búsqueda por novedad.

1. INTRODUCCIÓN

Las herramientas para predecir el comportamiento de los incendios forestales son de gran interés para la toma de decisiones en el control de incendios, con el fin de mitigar sus consecuencias. Existen diversos simuladores, los cuales utilizan modelos de propagación cuyo objetivo es predecir la evolución de la línea de fuego a lo largo de un período de tiempo, representada por mapas que indican el área del terreno alcanzada por el fuego en cada paso de predicción. Desde el punto de vista computacional, este problema de predicción es desafiante debido a la complejidad de los modelos utilizados y a las fuentes de incertidumbre involucradas en los datos de entrada. Este último aspecto es de gran importancia dado que las limitaciones para proveer parámetros de entrada correctos para el modelo conducen a errores en la predicción. Esta incertidumbre se debe a la dificultad o imposibilidad de obtener los valores de las variables involucradas. Para enfrentar esta incertidumbre, se han propuesto métodos que combinan resultados de múltiples simulaciones para analizar la tendencia, y así mitigar los efectos negativos de la incertidumbre. Recientemente, se han desarrollado dos

sistemas con este enfoque: ESSIM-EA [1, 2] y ESSIM-DE [3]. Ambos combinan una etapa de Optimización, para obtener escenarios cuya simulación permita obtener buenas predicciones, con una etapa de Análisis Estadístico, que logra la agregación de mapas obtenidos durante la optimización, para así poder predecir el comportamiento de la línea de fuego. Para manejar la alta carga de cómputo, ambos se basan en un *modelo de islas* [4], utilizando una doble jerarquía de procesos que les permite abarcar mejor el espacio de búsqueda y acelerar la optimización. Adicionalmente, y con el fin de mejorar la calidad y la eficiencia de estos métodos, se ha trabajado en incorporar a estos sistemas distintos mecanismos de *sintonización*. Las estrategias de sintonización permiten calibrar algún aspecto crítico, cuello de botella o factor limitante de la aplicación para mejorar su desempeño, y estas pueden ser automáticas (cuando las técnicas están incorporadas en la aplicación de forma transparente) y dinámicas (los ajustes ocurren durante la ejecución) [5].

A pesar de que los enfoques mencionados han obtenido mejoras respecto a métodos anteriores, aún tienen algunas limitaciones. Las metaheurísticas de estos enfoques utilizan una función de aptitud o *fitness* para evaluar la calidad de las soluciones. Esta función se usa para evaluar la calidad de los individuos (soluciones encontradas durante la búsqueda) y así guiar la búsqueda de manera que las poblaciones sucesivas contengan individuos con mejor *fitness* a lo largo de las iteraciones, con el objetivo de que la población eventualmente converja a una única solución. Mediante este mecanismo, a lo largo del tiempo aparecen genotipos (codificación de los individuos) cada vez más similares entre sí, por lo que, en este caso, la población que fue evolucionada para realizar las predicciones en cada intervalo de tiempo puede consistir en un conjunto de escenarios muy cercanos unos a otros, lo que limita el aporte de estas soluciones a la reducción de incertidumbre.

Además, se dejan de lado soluciones que pueden estar genotípicamente alejadas en el espacio de búsqueda, pero que aun así pueden tener valores de aptitud aceptables que contribuyan a la predicción. Estas limitaciones llevan a considerar la selección de otros enfoques de búsqueda que puedan producir mejoras en la calidad de las predicciones.

En base a este análisis, se observó la posibilidad de utilizar un paradigma alternativo al de las metaheurísticas tradicionales: la *búsqueda basada en novedad* o *Novelty Search* (NS) [6–8]. Este es un paradigma de búsqueda que ignora el objetivo como guía para la exploración y, en su lugar, recompensa a soluciones candidatas que presentan comportamientos novedosos (diferentes a los anteriormente descubiertos), con el fin de maximizar la exploración y evitar óptimos locales. Se ha aplicado con buenos resultados a múltiples problemas de diversas áreas [6, 7], [9–11]. La búsqueda basada en novedad es una alternativa prometedora frente a las limitaciones de las metaheurísticas aplicadas previamente al problema, debido a que, en este paradigma alternativo, la búsqueda es dirigida por una caracterización del comportamiento de los individuos, cambiando así el *paisaje de la función de fitness* (*fitness landscape*), previniendo de manera directa los problemas de convergencia prematura y estancamiento. Por último, en la literatura existen múltiples enfoques híbridos que combinan *fitness* y novedad y se ha demostrado que son efectivos para resolver problemas prácticos [7, 9, 10], [12–15]. Estas contribuciones indican que son muchas las alternativas a explorar en el área de NS, y por lo tanto las oportunidades de adaptabilidad de este tipo de algoritmos a distintos problemas.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Los sistemas de predicción existentes han obtenido resultados satisfactorios para la reducción de incertidumbre en el problema de aplicación. En particular, ESSIM-EA demostró obtener predicciones de buena calidad, mientras que ESSIM-DE redujo significativamente los tiempos de respuesta, pero sin obtener mejoras en calidad. Por esta razón, posteriormente se ha trabajado en mejorar el rendimiento del método ESSIM-DE mediante estrategias denominadas de sintonización [16]. Se desarrollaron dos de estas métricas [17, 18], ambas orientadas a mitigar los problemas de convergencia prematura y estancamiento de la población presentes en el caso de aplicación del algoritmo, obteniendo mejoras en calidad y tiempos de respuesta.

Dadas las características del problema y los resultados previos, se espera que un algoritmo que implemente NS pueda obtener buenos resultados en calidad y, dependiendo del algoritmo elegido y sus posibilidades de paralelización, también mejoras en eficiencia. El enfoque NS es adecuado para este problema, por un lado, porque resuelve por diseño los problemas propios de algoritmos explotativos; por otro lado, porque presenta múltiples oportunidades de paralelización y de hibridación con enfoques basados en *fitness*. También es posible adaptar el comportamiento del algoritmo de acuerdo a las características del problema. Respecto a la posibilidad de aplicación de mecanismos de sintonización, en el caso de NS, también es factible suponer que un proceso de sintonización sería capaz de contribuir predicciones más precisas y en menor tiempo para el caso propuesto; por ejemplo, implementando la modificación, de forma automática y dinámica, de alguno de los aspectos antes mencionados. La presente línea de investigación plantea la implementación de una metaheurística adaptada al paradigma NS,

comenzando con una versión simple para luego poder implementar mejoras iterativas en base a las posibilidades previamente descritas. En una primera versión se paralelizará solo el cómputo de las evaluaciones de individuos, mientras que a futuro se podrán explorar métodos como el modelo de islas para este nuevo enfoque.

3. OBJETIVOS TEÓRICOS Y EXPERIMENTALES

El objetivo general de esta investigación consiste en la reducción de incertidumbre en sistemas de predicción de fenómenos de propagación mediante un enfoque paralelo de búsqueda por novedad, aplicado al caso de incendios forestales. Se espera que este nuevo enfoque pueda obtener resultados experimentales de calidad comparable o superior a los métodos existentes. Posteriormente, mediante la incorporación de métodos adicionales de paralelización, se esperan posibles mejoras tanto en calidad como en eficiencia, respecto a la primera versión. A nivel teórico, se espera que los resultados experimentales provean nuevo conocimiento que permita caracterizar mejor el problema, mediante el análisis de las predicciones basadas en NS y su comparación con resultados de métodos anteriores.

A largo plazo, se plantean objetivos que pueden obtener mejoras adicionales. Por un lado, extender el método desarrollado mediante estrategias híbridas de búsqueda que combinen novedad y *fitness*. Por otro lado, diseñar métodos de sintonización que permitan variar los parámetros de dichas estrategias híbridas con el fin de proveer mayor adaptabilidad al algoritmo frente a diversas instancias del problema.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo para esta investigación se encuentra conformado por el Dr. Jan Strappa,

quien se incorporó al laboratorio en 2021 con una beca postdoctoral de CONICET, dirigida por la Dra. Paola Caymes Scutari y codirigida por el Dr. Germán Bianchini. En 2020 se defendieron dos tesis doctorales relacionadas con esta línea de investigación, desarrolladas por la Dra. Laura Tardivo (bajo la dirección de la Dra. Paola Caymes Scutari) y el Dr. Miguel Méndez Garabetti (bajo la dirección del Dr. Germán Bianchini y la codirección de la Dra. Paola Caymes Scutari).

El grupo de trabajo siempre está abierto a la incorporación de nuevos integrantes (de grado o postgrado) que deseen familiarizarse con las temáticas con las que se trabaja dentro del mismo.

5. BIBLIOGRAFÍA

- [1] M. Méndez Garabetti, G. Bianchini, M. L. Tardivo, and P. Caymes Scutari, "Comparative Analysis of Performance and Quality of Prediction Between ESS and ESS-IM," *Electronic Notes in Theoretical Computer Science*, vol. 314, pp. 45–60, Jun. 2015.
- [2] M. Méndez Garabetti, G. Bianchini, P. Caymes Scutari, M. L. Tardivo, and V. Gil Costa, "ESSIM-EA applied to Wildfire Prediction using Heterogeneous Configuration for Evolutionary Parameters," *XXIII Congreso Argentino de Ciencias de la Computación*, p. 10, 2017.
- [3] M. L. Tardivo, P. Caymes Scutari, G. Bianchini, and M. Méndez Garabetti, "Hierarchical parallel model for improving performance on differential evolution: Hierarchical parallel model for improving performance on differential evolution," *Concurrency and Computation: Practice and Experience*, vol. 29, no. 10, p. e4087, May 2017.
- [4] E. G. Talbi, *Metaheuristics: From Design to Implementation*. 2009.

- [5] P. Caymes Scutari, G. Bianchini, A. Sikora, and T. Margalef, “Environment for automatic development and tuning of parallel applications,” in *2016 International Conference on High Performance Computing & Simulation (HPCS)*, 2016, pp. 743–750.
- [6] J. Lehman and K. O. Stanley, “Abandoning Objectives: Evolution Through the Search for Novelty Alone,” *Evolutionary Computation*, vol. 19, no. 2, pp. 189–223, Jun. 2011.
- [7] J. Lehman and K. O. Stanley, “Exploiting Open-Endedness to Solve Problems Through the Search for Novelty,” *Artificial Life*, p. 8, Jan. 2008.
- [8] J. Lehman and K. O. Stanley, “Evolvability Is Inevitable: Increasing Evolvability without the Pressure to Adapt,” *PLoS ONE*, vol. 8, no. 4, pp. 2–10, 2013.
- [9] J. Gomes, P. Urbano, and A. L. Christensen, “Evolution of swarm robotics systems with novelty search,” *Swarm Intelligence*, vol. 7, nos. 2-3, pp. 115–144, Sep. 2013.
- [10] P. Krčah, “Solving Deceptive Tasks in Robot Body-Brain Co-evolution by Searching for Behavioral Novelty,” in *Advances in Robotics and Virtual Reality*, vol. 26, J. Kacprzyk, L. C. Jain, T. Gulrez, and A. E. Hassanien, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 167–186.
- [11] C. Ollion and S. Doncieux, “Why and how to measure exploration in behavioral space,” in *Proceedings of the 13th annual conference on Genetic and evolutionary computation - GECCO '11*, 2011, p. 267.
- [12] J. Gomes, P. Mariano, and A. L. Christensen, “Devising Effective Novelty Search Algorithms: A Comprehensive Empirical Study,” in *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*, 2015, pp. 943–950.
- [13] J.-B. Mouret and S. Doncieux, “Encouraging Behavioral Diversity in Evolutionary Robotics: An Empirical Study,” *Evolutionary Computation*, vol. 20, no. 1, pp. 91–133, Mar. 2012.
- [14] J. K. Pugh, L. B. Soros, P. A. Szerlip, and K. O. Stanley, “Confronting the Challenge of Quality Diversity,” in *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*, 2015, pp. 967–974.
- [15] J.-B. Mouret and J. Clune, “Illuminating search spaces by mapping elites,” *arXiv:1504.04909 [cs, q-bio]*, Apr. 2015.
- [16] K. Naono, K. Teranishi, J. Cavazos, and R. Suda, Eds., *Software Automatic Tuning*. New York, NY: Springer New York, 2010.
- [17] M. L. Tardivo, P. Caymes Scutari, M. Méndez Garabetti, and G. Bianchini, “Optimization for an Uncertainty Reduction Method Applied to Forest Fires Spread Prediction,” in *Computer Science 2017*, vol. 790, A. E. De Giusti, Ed. Cham: Springer International Publishing, 2018, pp. 13–23.
- [18] M. L. Tardivo, P. Caymes Scutari, G. Bianchini, and M. Méndez Garabetti, “Sintonización Dinámica del Método Paralelo de Predicción de Incendios Forestales ESSIM-DE,” *XXV Congreso Argentino de Ciencias de la Computación*, p. 10, 2019.

Técnicas de modelado y simulación en sistemas de HPC y salud.

Diego Encinas^{1,2}, Jimena Jara¹, Román Bond¹, Daniel Rosatto¹, Lucas Maccallini¹, Mauro Gomez¹, Federico Montes de Oca¹, Adriana Gaudiani³, Martín Morales^{1,4}

¹Proyecto de Investigación SimHPC - Programa TICAPPS - Instituto de Ingeniería y Agronomía - UNAJ

²Instituto de Investigación en Informática LIDI (III-LIDI) - Facultad de Informática - UNLP - Centro Asociado CIC

³Área de Computación - Instituto de Ciencias - UNGS

⁴Centro CodApli - Facultad Regional La Plata - UTN

dencinas@unaj.edu.ar, elchejime@gmail.com, rbond@unaj.edu.ar, drosatto@unaj.edu.ar, lucas.maccallini@gmail.com, mauro.f.gmz@gmail.com, federicomdo97@gmail.com, agaudiani@ungs.edu.ar, martin.morales@unaj.edu.ar

Resumen

El objetivo de esta línea de investigación es el estudio de diferentes técnicas de modelado y simulación para entornos de Cómputo en Altas Prestaciones (HPC, High Performance Computing). El enfoque del estudio es la obtención de herramientas que permitan predecir la eficiencia del sistema ante posibles escenarios y reconfigurar el sistema físico. Además, se analizan los diferentes componentes del sistema que pueden influir en las prestaciones significativamente y pueden llegar a modelarse y/o reconfigurarse.

Por otra parte, se ha llevado a cabo el desarrollo de un simulador para entornos en el área de salud, en el sector de emergencias hospitalarias y de propagación de enfermedades.

Palabras clave: *Arquitecturas Multiprocesador. Simulación. Sistema E/S paralela. Modelado y Simulación basado en agentes (Agent-Based Modeling and Simulation, ABMS). Cloud Computing. CloudSim. Simulación y Salud.*

Contexto

Se presenta una línea de Investigación que es parte del Proyecto de Investigación “Simulación, Computación de Altas Prestaciones (HPC) y

optimización de aplicaciones sociales – SimHPC” de la Universidad Nacional Arturo Jauretche (UNAJ), acreditado por resolución interna 183/21. Además, el proyecto aporta al Programa “Tecnologías de la información y la comunicación (TIC) en aplicaciones de interés social” – TICAPPS de la UNAJ.

En el tema existe un convenio de colaboración en actividades de Investigación y Postgrado con el Instituto de Investigación en Informática – LIDI de la Universidad Nacional de La Plata y el Área de Computación del Instituto de Ciencias de la Universidad Nacional de General Sarmiento.

Introducción

El crecimiento sostenido en la demanda del poder de cómputo remarca la necesidad de sistemas con enfoques de paralelización masiva y cómputo de alta performance (HPC, High Performance Computing) [1]. Los clusters se han convertido en uno de los enfoques principales para lograr paralelismo a bajo costo. Una noción extendida lo constituye la utilización de grid computing y más recientemente cloud computing. Independientemente de la solución, estos sistemas constan de un gran número de componentes incluyendo nodos de

procesamiento, bancos de memoria, discos, entre otros.

En cuanto a las herramientas de simulación para Cloud Computing, CloudSim es un framework desarrollado en Java que provee las APIs necesarias para que el usuario genere una simulación de un Data Center funcionando como servidor de nube capaz de simular la ejecución de CloudLets.

Por otra parte, los servicios de urgencias hospitalarias son considerados como una de las unidades del sistema sanitario de mayor complejidad y fluidez, lo que unido a la variabilidad de su actividad da lugar a que su gestión operativa sea una tarea muy complicada [2]. Es por ello que resultaría de mucha utilidad para sus responsables disponer de un sistema de ayuda a la toma de decisiones (Decision Support System-DSS) tan flexible como un simulador, que permitiría tomar medidas disponiendo de información suficiente sobre las alternativas posibles.

Sistemas de E/S Paralela

Las exigencias en los sistemas de E/S paralelos se han incrementado debido al aumento en número, velocidad y potencia de las unidades de procesamiento en los clusters. También las aplicaciones científicas que utilizan cómputo de altas prestaciones acrecientan estos requerimientos.

En muchos casos, el cuello de botella de los sistemas paralelos es la E/S a causa de las exigencias que debe afrontar [3]. La E/S Paralela es esencial para emparejar el avance de las arquitecturas de los procesadores y el rápido crecimiento de la capacidad computacional. Aunque la arquitectura jerárquica de memoria multinivel puede evitar grandes pérdidas de prestaciones debido a los retardos de acceso a disco, la capacidad de memoria es limitada. Además, como la capacidad computacional aumentará, la disponibilidad de memoria por core decrecerá, especialmente si la escala de los sistemas de HPC se proyecta a millones de cores o más. Varias simulaciones científicas y de ingeniería de áreas críticas de investigación, tales como la nanotecnología, astrofísica, clima y energía física están convirtiéndose en aplicaciones intensivas de datos. Para poder disminuir la brecha entre CPUs-E/S se deben

identificar los factores que influyen en las prestaciones y proponer nuevas soluciones [4] [5].

En el área de tolerancia a fallas en sistemas de cómputo de alta prestaciones se puede notar la importancia de la unidad de E/S en las arquitecturas paralelas como un punto a mejorar para lograr cubrir las exigencias de las aplicaciones que utilizan HPC. Una manera de llevar a cabo este trabajo es utilizar técnicas de simulación para evaluar el efecto de los cambios de los factores con mayores influencias en las prestaciones del sistema de E/S paralelo.

Se puede disminuir la complejidad y la probabilidad de errores en la generación de sistemas híbridos desarrollando una simulación específica de éstos utilizando diferentes frameworks [6] [7] [8].

Las aplicaciones científicas con un uso intensivo de datos utilizan software de E/S paralelo para acceder a archivos. Contar con una herramienta que permita predecir el comportamiento de este tipo de aplicaciones en HPC es de gran utilidad para los desarrolladores de aplicaciones paralelas como para administradores de centros de cómputo. Por otro lado, ABMS ha sido utilizado para modelar problemas y sistemas complejos en diversas áreas de la ciencia.

Evaluar las prestaciones del subsistema de E/S con diferentes configuraciones y la misma aplicación, permite adaptar la configuración de E/S teniendo en cuenta el patrón de acceso de la aplicación. Pero también puede ser una gran ventaja analizar las necesidades de las aplicaciones antes de configurar el sistema físico. Una manera de predecir el comportamiento de las aplicaciones en el sistema de cómputo, ante distintas configuraciones, es utilizando técnicas de modelado y simulación.

Se está desarrollando modelos e implementando una simulación de la arquitectura de E/S paralela, por medio de técnicas de simulación basadas en agentes o Sistemas Multi-Agente, (MAS-MultiAgentSystems), para evaluar el efecto de dimensionar el sistema de E/S o cambiar componentes como la red de almacenamiento, dispositivos de E/S, entre otros [9].

Simulación de arquitecturas de Cloud Computing

CloudSim [10] [11] es un Framework de simulación generalizado y extensible que permite el modelado y la simulación de diferentes infraestructuras y servicios de aplicaciones de Cloud Computing. Un ejemplo de utilización es la simulación de muchos centros de datos.

Su arquitectura consiste en entidades específicas que se representan como clases Java que pueden ser heredadas o instanciadas. Estas clases representan centros de datos, hosts físicos, máquinas virtuales, servicios a ejecutar en los centros de datos y servicios en la nube de usuarios [12] [13]. Además, CloudSim soporta la inserción dinámica de los elementos de simulación y proporciona aplicaciones de paso de mensajes y la topología de la red del centro de datos.

La versatilidad de CloudSim es la principal ventaja del sistema. La integración de nuevos parámetros y conceptos de la simulación es implementada desde abstracciones preestablecidas convenientemente por los autores. Las abstracciones principales son SimEvent [14], SimEntity [15], DataCenterCharacteristics y Vm.

El aporte de un desarrollo de nuevos actores al componente Vm (Virtual Machine) posibilita que por medio de simulación se obtengan métricas de entrada/salida. Las estadísticas que aporta la nueva implementación de Vm dan soporte a un espacio de memoria ram en tiempos de simulación. Dicho espacio de memoria principal está controlado por otro componente que es capaz de procesar instrucciones guardadas en el espacio de memoria sintético denominado RamEntity. Si las instrucciones se guardan lógicamente, la nueva versión de Vm es capaz de administrar procesos en la nueva capa de ejecución. La idea general de la implementación consiste en la creación de una nueva SimEntity y sus agregaciones necesarias para manipular el espacio de memoria proporcionada por RamEntity, en tiempos de simulación. De tal forma que cada celda de RamEntity no sólo guarda datos, sino que también es posible que mantenga objetos interpretables como eventos para la nueva SimEntity que se denomina InitEntity.

Lo destacable es que el entorno de la cola de procesos es en el espacio de memoria de RamEntity y que cada instrucción que compone a un código objeto del proceso es en realidad un evento con todos sus parámetros.

Simulación y Salud

Al analizar las necesidades presentes en los servicios de salud, se obtiene que dentro de una sala de urgencias se encuentre una gran diversidad de escenarios posibles donde cada uno de estos puede afectar a resultados sensibles como, por ejemplo, la tasa de mortalidad de personas. Para solucionar este problema se desarrolló un simulador con el objetivo de ser una herramienta capaz de recrear una gran cantidad de escenarios y así poder tomar decisiones rápidas.

El simulador desarrollado se centra en la prevalencia puntual de infecciones intrahospitalarias [16] en una sala de urgencias y cómo la afectan distintos factores relacionados con la gestión hospitalaria.

Para llevar a cabo el modelado del simulador se utilizó el paradigma de Modelado y Simulación basado en Agentes (ABMS). El framework utilizado es Repast Symphony [17], una herramienta especializada en ABMS, la cual provee una serie de ventajas respecto al modelado e implementación de agentes, así como también la coordinación concurrente de los agentes.

De esta manera, se clasificaron distintos agentes intervinientes en el ambiente de una sala de urgencias, como los pacientes y médicos, entre otros. Dichos agentes definen su comportamiento mediante máquinas de estado, las cuales determinan las acciones correspondientes tanto a la atención hospitalaria como también al estado de salud respecto de una enfermedad intrahospitalaria específica.

La propagación de la infección intrahospitalaria estará modelada mediante interacciones entre estos agentes, por ejemplo, uno de los focos de infección más importante de las salas de urgencias son las salas de espera. La forma más efectiva para calibrar el simulador se da mediante el grado de interacción de agentes, ajustando así la tasa de transmisión de la enfermedad en cuestión.

Una vez calibrado el simulador, se pueden obtener resultados. El trabajador de la salud dispone de diversos parámetros para configurar la simulación, por ejemplo, la cantidad de pacientes, la cantidad de camas disponibles, la cantidad de insumos hospitalarios, etc. Al realizar distintas ejecuciones, se pueden obtener resultados analizando distintos posibles cuellos de botella, configurando la cantidad de médicos clínicos disponibles, la cantidad de recepcionistas encargados de la admisión, el triage, entre otros. De esta manera, es posible la toma de decisiones respecto a la asignación de recursos y personal para agilizar la estancia hospitalaria de los pacientes y evitar los contagios producidos por sus interacciones.

Además, se analiza y trabaja la optimización y paralelización del simulador con los ambientes Repast HPC para clusters y Flame para ejecución sobre GPU, obteniendo mejor performance en términos de tiempos.

Líneas de Investigación, Desarrollo e Innovación

Temas de Estudio e Investigación

- Arquitecturas multiprocesador para procesamiento paralelo: multiprocesador de memoria compartida, multiprocesador on-chip de memoria distribuida. Multicore, Clusters, Clusters de multicore. Grid. Cloud.
- Arquitectura de E/S paralela considerando el software, hardware, comunicaciones entre módulos y dispositivos de almacenamiento.
- Nuevos aportes de desarrollos que mejoren los modelos de simulaciones con CloudSim para el análisis de la performance en sistemas de arquitecturas de software de Cloud Computing.
- Modelado y simulación para la administración de sistemas de salud.

Resultados y Objetivos

Investigación experimental

- Diseño y desarrollo de modelos mediante técnicas de Modelado y simulación basada en agentes (ABMS) para analizar el

comportamiento de las distintas capas de la pila de software de E/S.

- Análisis, modelado e implementación de las operaciones típicas de E/S: read, write, open, close, flush.
- Análisis y modelado de librerías de archivos para aplicaciones que utilizan cómputo de altas prestaciones-HPC. Se ha utilizado Amazon Web Services para creación de cluster virtuales y obtener métricas de la pila de software de E/S.
- Incorporación de tiempos de entrenamiento, obtenidos en AWS, en el simulador. Con esto se logra una salida más detallada y un método para validar tiempos y métricas del simulador con AWS.
- Implementación de comandos para ejecutar desde command center en NetLogo. Con esto se logró sintetizar el benchmark IOR correspondiente a la capa de aplicación de la pila de E/S, logrando introducir nuevos parámetros como tamaño de archivo y cantidad de nodos de E/S (metadata server y data server). De ejecutar este comando, se obtiene una nueva salida similar a la del benchmark IOR [18].
- Obtención de un método de desarrollo de nuevos actores genéricos CloudSim que mejoran el modelado y la producción de estadísticas virtuales.
- Implementación de la entidad InitEntity que procesa instrucciones en el espacio de memoria de las máquinas virtuales. Se vinculan exitosamente los tiempos de procesamiento de un cloudlet y las instrucciones en las máquinas virtuales [19].
- Contraste de una simulación de un cluster en la nube y uno idéntico desplegado en un sistema de cloud computing público [20].
- Modelado de infraestructuras de sistemas de salud.
- Implementación de un simulador para analizar el contagio de enfermedades intrahospitalarias [21].
- Paralelización en la ejecución de simuladores.
- Implementación de un simulador para analizar la propagación de enfermedades [22].

Formación de Recursos Humanos

Dentro de la temática de la línea de I/D se participa en el dictado de la carrera de Ingeniería en Informática de la UNAJ. También aportan trabajos de alumnos de las materias Redes de Computadoras 2 y Programación en Tiempo Real. Por otro lado, algunos integrantes participan en el dictado de la Diplomatura en Ciencia de Datos de la UNAJ.

Durante 2021 se han realizado publicaciones nacionales e internacionales. Además, se encuentran en desarrollo y concluidas varias Prácticas Profesionales Supervisadas (PPS) con las que concluyen sus estudios los alumnos de Ingeniería en Informática.

En esta línea de I/D existe cooperación a nivel nacional e internacional. Hay dos investigadores realizando estudios de postgrado, 1 becario EVC CIN, un becario de Iniciación a la Investigación UNAJ y 2 alumnos avanzados de grado colaborando en las tareas.

Referencias

1. Grama A, Gupta A, Karypis G, Kumar V. "Introduction to parallel computing". Second Edition. Pearson Addison Wesley, 2003.
2. R. Galeano, C. Villalba, D. Rexachs, E. Luque. Agent-Based Model to Simulate Outpatient's Consultations at the "Hospital de Clínicas". The Eighth International Conference on Advances in System Simulation (SIMUL 2016). 1:46-51
3. H Hennessy, J. L., Patterson, and D. A., Computer Architecture, Fourth Edition: A Quantitative Approach. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2006.
4. J. M. May, Parallel I/O for high performance computing. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001.
5. V. Balaji, Earth system modelling – Volume 4. IO and Postprocessing. Springer, 2013.
6. D. Encinas, Utilización de un reloj global para el modelado de un ambiente simulado distribuido. XVIII Congreso Argentino de Ciencias de la Computación. 2012
7. D. Encinas, Simulación de una red CAN para dimensionar las comunicaciones de una IMU. VII Congreso Argentino de Tecnología Espacial. 2013.
8. D. Black, SystemC: From the Ground Up. Second Edition, Springer, 2010.
9. D. Encinas et al., Modeling I/O System in HPC: An ABMS Approach. The Seventh International Conference on Advances in System Simulation (SIMUL), ISBN: 978-1-61208-442-8, 2015.
10. R. Calheiros, R. Ranjan, A. Beloglazov, C. De Rose and R. Buyya "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms" Published online 24 August 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/spe.995.
11. <http://www.cloudbus.org/cloudsim> 2018.
12. Hamza Ouarnoughi, Jalil Boukhobza, Frank Singhoff, Stephane Rubini, Erwann Kassis. "Considering I/O Processing in CloudSim for Performance and Energy Evaluation". OpenStack Cloud Software: Open source software for building private and public clouds. © Springer International Publishing AG 2016 M. Tauber et al. (Eds.): ISC High Performance Workshops 2016, LNCS 9945, pp. 591–603, 2016. DOI: 10.1007/978-3-319-46079-6_40.
13. Kushang Parikh, Nagesh Hawanna, Haleema. P.K, Jayasubalakshmi.R and N.Ch.S.N.Iyengar. School of Computing Science and engineering Vellore Institute of Technology, Tamil Nadu, "Virtual Machine Allocation Policy in Cloud Computing Using CloudSim in Java." 2015.
14. www.icsa.inf.ed.ac.uk/research/groups/hase/simj ava 2018.
15. F. Howell, R Mc Nab. A discrete event simulation library for java. International Conference on Web-Based Modeling and Simulation. 1998.
16. Prevalencia puntual de infección nosocomial, disponible en internet: https://www.researchgate.net/publication/242363760_Prevalencia_puntual_de_infeccion_nosocomial. Fecha: 29/03/2020
17. Repast Symphony Frequently Asked Questions, disponible en internet: <https://repast.github.io/docs/RepastFAQ/RepastFAQ.html>. Fecha: 29/03/2020
18. D. Encinas, S. Mendez, M. Naiouf, A. De Giusti, D. Rexachs del Rosario, and E. Luque, An Agent-Based Model for Analyzing the HPC Input/Output System International journal on advances in systems and measurements vol. 13, num. 3 & 4, págs. 192-202, 2020.
19. D. Rosatto, R. Bond, M. Belizán, M. Morales, D. Encinas. Modelado y simulación de arquitecturas de Cloud Computing con CloudSim: comunicación entre entidades. XXIII Congreso Argentino de Ciencias de la Computación. ISBN 978-950-34-1539-9. 2017
20. T. Rosales, J. Spinelli, M. Di Nardo, R. Bond, D. Rosatto, D. Encinas, F. Romero. Análisis de una plataforma de simulación para Cloud Computing. Un caso de estudio. XXVI Congreso Argentino de Ciencias de la Computación. 2020
21. L. Maccallini, D. O. Encinas, and F. Romero. "An Approach to the Modeling and Simulation of Intra-Hospital Diseases". Journal of computer science and technology (ISSN 1666-6038), vol. 21, num. 2, págs. 157-169, doi. 10.24215/16666038.21.e14, 2021.
22. J. Baez, A. Barreto, B. Galarza, M. Morales, D. Encinas. Simulación para estimar propagación de enfermedades. 8º Congreso Nacional de Ingeniería Informática/Sistemas de Información (CoNaIISI 2020). San Francisco, Córdoba. 2020

Cloud Computing, IaaS privados y públicos para el análisis y modelado de sistemas.

Diego Encinas^{1,2}, Brian Galarza¹, Román Bond¹, Gonzalo Zaccardi¹, Nicolás Benquerença Mendes¹, Jorge Osio¹, David Duarte¹, Martín Morales^{1,3}

¹Proyecto de Investigación SimHPC - Programa TICAPPS - Instituto de Ingeniería y Agronomía - UNAJ

²Instituto de Investigación en Informática LIDI (III-LIDI) - Facultad de Informática - UNLP – Centro Asociado CIC

³Centro CodApli - Facultad Regional La Plata - UTN

dencinas@unaj.edu.ar, bgalarza@unaj.edu.ar, rbond@unaj.edu.ar, gzaccardi@unaj.edu.ar, nicobenquerenca@yahoo.com.ar, josio@unaj.edu.ar, davito.duarte.22@gmail.com, martin.morales@unaj.edu.ar

Resumen

El objetivo de esta línea de investigación es el estudio del rendimiento de las arquitecturas tipo cloud a través del despliegue de IaaS y utilización de IaaS públicos, en particular en el área de cómputo paralelo de altas prestaciones (HPC). Enfocando en la obtención de herramientas que permitan predecir la eficiencia del sistema ante posibles escenarios. Analizando los diferentes componentes del sistema que pueden influir en las prestaciones significativamente, especialmente la entrada/salida y las comunicaciones.

Palabras clave: *Cloud Computing. OpenStack. Sistemas de Archivos en clústers. Redes definidas por Software.*

Contexto

Se presenta una línea de Investigación que es parte del Proyecto de Investigación “Simulación, Computación de Altas Prestaciones (HPC) y optimización de aplicaciones sociales – SimHPC” de la Universidad Nacional Arturo Jauretche (UNAJ), acreditado por resolución interna 183/21. Además, el proyecto colabora con el Programa “Tecnologías de la información y la comunicación (TIC) en aplicaciones de interés social” – TICAPPS de la UNAJ.

Introducción

Cloud Computing es un paradigma que ha estado en constante crecimiento. Cada vez más compañías y grupos de investigación trabajan en conjunto con el fin de explotar las oportunidades ofrecidas por el mismo [1]. Dicho paradigma ofrece muchas ventajas, tales como el bajo costo de implementación, ya que no se necesitan

computadoras de última tecnología debido a que éstas trabajan conjuntamente (Clustering) con la posibilidad de escalar horizontalmente de manera sencilla. Además, hay software Open Source disponible para los nodos en el clúster como las infraestructuras Eucalyptus, OpenNebula, CloudStack u OpenStack integradas con GNU/Linux y compatibles, por ejemplo, con Amazon WebServices.

Despliegue de IaaS

Las comunicaciones en Cloud Computing son una parte fundamental del paradigma que consisten en utilizar distintos nodos y lograr hacerlos funcionar conjuntamente.

Para lograr una comunicación sincronizada entre estos nodos se propone utilizar OpenStack [2] como también OpenNebula [3].

OpenStack es una plataforma de tecnología open source que utiliza recursos virtuales agrupados para diseñar y gestionar nubes privadas y públicas a través de múltiples servicios que, de manera coordinada, cumplen diferentes propósitos para lograr el correcto funcionamiento de, por ejemplo, una "Infraestructure as a service" (IaaS). Algunos de los servicios ofrecidos por OpenStack son: hypervisor (Nova), autenticación (Keystone), Imágenes (Glance), Dashboard (Horizon), Networking (Neutron) y block storage (Cinder). Según las necesidades se pueden requerir de ciertos servicios u otros. La Arquitectura básicamente consiste en dos tipos de nodos: "Compute Node" y "Controller Node". Se llaman Compute Node a todos aquellos que se encargan del procesamiento de servicios específicos mientras que Controller Node es aquel que comunica a cada uno de los anteriores [4] [5] [6].

Fuel es una herramienta open source desarrollada por Mirantis en la cual se ejecuta un script que permite configurar, de manera más amigable respecto a OpenStack, los recursos que se desean otorgar a la infraestructura, como la cantidad de nodos, los núcleos de procesador, la memoria RAM, entre otros [7].

Fuel trabaja con un nodo master el cual es el encargado de controlar a los nodos slaves que contendrán la infraestructura OpenStack. Es decir, desde el nodo Fuel Master se indican qué paquetes se van a instalar en cada nodo slave (Glance, Nova-Compute, Keystone, etc.) para luego en los slaves tener armados los nodos compute y controller, sin necesidad de realizar configuraciones manuales en cada uno de los mismos.

OpenNebula es un software de código abierto que permite el despliegue de IaaS. Busca reducir la complejidad generada por OpenStack y ofrece soporte con hipervisores tales como KVM y VMware vCenter.

OpenNebula clasifica a los nodos en dos tipos, Front – end los cuales entran en contacto con los usuarios y a su vez se comunican con los nodos de la infraestructura en los cuales se lanzarán las instancias y los nodos virtualizados los que a su vez deben contar con los paquetes correspondientes de storage, autenticación y networking para poder funcionar correctamente.

La implementación de estas infraestructuras ofrece ventajas en las cuales los clústeres virtualizados trabajan en conjunto ofreciendo un buen rendimiento a bajos costos y con posibilidad de escalabilidad al poder agregar mayor cantidad de nodos para procesamiento de manera sencilla.

Sistemas de Archivos Paralelos en clústeres

Amazon Webservices [8], mediante el servicio EC2 (Amazon Elastic Compute Cloud) permite desplegar clústeres virtuales mediante instancias de VMs y almacenamiento para las mismas. Mediante este servicio y el correspondiente clúster conformado, se utiliza un sistema de archivos paralelo (PVFS2) que permite la gestión de datos particionados y distribuidos en los distintos nodos, mediante múltiples tareas de una aplicación ejecutada sobre el clúster.

PVFS2 utiliza una estructura cliente-servidor. Dependiendo de su rol, existen tres tipos de nodos dentro de un clúster con este sistema de archivos: servidores de datos, servidores de metadatos y clientes, en donde cada uno de los nodos puede cumplir los tres roles.

PVFS2 [9] contiene una herramienta interna que permite volcar información en logs durante la ejecución de tareas dentro del sistema de archivos. Por cada uno de los nodos, el administrador del clúster puede obtener información de depuración (GOSSIP) conformado por registros de debug, de acceso, contadores de rendimiento y errores producidos durante la ejecución. Referido a los contadores de rendimiento, se obtienen mediante el software de monitoreo Atop. Esta herramienta permite obtener reportes de la actividad de los procesos y la utilización de los diferentes recursos del sistema (memoria, disco, CPU, red, etc) [10].

Asimismo, es posible especificar los datos a obtener en función de las distintas capas de PVFS2, del rol específico de cada nodo (cliente-servidor), así como también de otros factores relacionados con el funcionamiento interno y operaciones asociadas a la gestión de archivos.

Teniendo conocimiento del funcionamiento del código fuente correspondiente a PVFS2 y del modo en el

que realiza la escritura de registros en los logs, resulta factible realizar modificaciones que permitan obtener otro tipo de parámetros de interés a partir de esta herramienta.

Finalmente se propone el análisis, uso y configuración de distintas herramientas no invasivas para determinar la performance del sistema de archivos en clústeres virtuales.

Redes definidas por Software

Las redes definidas por software (SDN) son un paradigma de gestión y administración de redes por medio de software que permiten tener un control más flexible con respecto al control del tráfico de datos por medio de hardware ya que permite cambiar en tiempo real las normas y políticas establecidas en la red[11].

Mininet[12] es un emulador de redes SDN open source que permite generar tráfico artificial entre nodos virtuales de la red.

Se ha llevado a cabo el montaje de un laboratorio SDN utilizando Mininet junto con Amazon WebServices con el fin de analizar el comportamiento de este nuevo paradigma de redes de computadoras en un entorno de Cloud Computing.[13]

Líneas de Investigación, Desarrollo e Innovación

Temas de Estudio e Investigación

- Arquitecturas multiprocesador para procesamiento paralelo: multiprocesador de memoria compartida, multiprocesador on-chip de memoria distribuida. Multicore, Clusters, Clusters de multicore. Grid. Cloud.

- Plataformas de software para implementar y administrar Clouds públicos, privados e híbridos.
- Sistemas de Archivos Paralelos.

Resultados y Objetivos

Investigación experimental

- Implementación de un IaaS encargado de realizar operaciones en procesamiento paralelo aumentando la eficiencia y reduciendo los costes generados.
- Implementación de OpenStack Dashboard y de un sistema desarrollado para poder controlar/administrar de manera visual (web) y más básica cada uno de los servicios.
- Implementación de OpenNebula en un sistema con las mismas características que el implementado por OpenStack con el fin de poder realizar pruebas en entornos similares.
- Utilización de Fuel para administrar OpenStack como sistema de administración de nube (Cloud Computing) a partir de la infraestructura de 2 nodos compute y el controller [14].
- Análisis del rendimiento de un Cloud privado en la ejecución de instancias personalizadas.
- Ejecutar diferentes benchmarks en la infraestructura desplegada sobre OpenStack y OpenNebula para así comparar los resultados obtenidos de ambas infraestructuras y poder realizar un análisis del rendimiento en cada caso.
- Introducción al estudio e integración de sensores físicos y los servicios en la nube [15].

- Medición de servicios en la nube enfocados a IaaS y PaaS [16].
- Análisis y configuración de clústeres virtuales.
- Análisis y configuración de herramientas no invasivas para la obtención de métricas en las distintas capas de software de los sistemas de archivos paralelos.
- Utilización de otros sistemas de archivos paralelos como Lustre [17] y Beegfs [18] para obtener métricas en Metadataservidores [19] [20].

Formación de Recursos Humanos

Dentro de la temática de la línea de I/D se participa en el dictado de la carrera de Ingeniería Informática de la UNAJ. También aportan trabajos de alumnos de las materias Sistemas Operativos 1, Redes de Computadoras 2, Programación en Tiempo Real y Organización y Arquitecturas de Computadoras. Por otro lado algunos integrantes participan en el dictado de la Diplomatura en Ciencia de Datos de la UNAJ.

Durante 2021 se han realizado publicaciones nacionales. Además, se encuentran en desarrollo y concluidas varias Prácticas Profesionales Supervisadas (PPS) con las que concluyen sus estudios los alumnos de Ingeniería en Informática.

En esta línea de I/D existe cooperación a nivel nacional. Hay 4 investigadores realizando carreras de postgrado y alumnos avanzados de grado colaborando en las tareas.

Referencias

1. Kondo, D., Javadi, B., Malecot, P., Cappello, F., Anderson, D. P.: "Cost-benefit

- analysis of Cloud Computing versus desktop grids”. In: IPDPS '09 Proceedings. IEEE International Symposium on Parallel and Distributed Processing. Washington, USA (2009).
2. OpenStack Cloud Software: Open source software for building private and public clouds. <http://www.openstack.org>. Febrero 2015.
 3. OpenNebula. <https://opennebula.org/>. Febrero 2019
 4. Galarza, B.; Tuamá, C.; Zaccardi, G.; Encinas, D.; Morales, M. “Implementaciones de Cloud Computing y aplicaciones en el ámbito universitario”. 1° Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNaIISI 2013). Ciudad de Córdoba, Argentina.
 5. Zaccardi, G.; Galarza, B.; Encinas, D.; Morales, M. “Implementación de Cloud Computing utilizando OpenStack”. 2° Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNaIISI 2014). Ciudad de San Luis, Argentina.
 6. Galarza, B.; Zaccardi, G.; Encinas, D.; Morales, M. “Análisis de despliegue de una IaaS utilizando Openstack”. XXI Congreso Argentino de Ciencias de la Computación (CACIC 2015). Ciudad de Junín, Argentina.
 7. OpenStack Deployment Fuel. <https://www.mirantis.com/products/mirantis-openstack-software/openstack-deployment-fuel/>. Febrero 2016
 8. Amazon Web Services (AWS)-Cloud Computing Services. <https://aws.amazon.com> Marzo 2019
 9. T. PVFS2, “PVFS 2 File System Semantics Document,” tech. rep., PVFS Development Team, 2015
 10. Atop Tool. <https://www.atoptool.nl/index.php>
 11. Kreutz, D., Ramos, F. M. V., Esteves Verissimo, P., Esteve Rothenberg, C., Azodolmolky, S., & Uhlig, S. (2015). Software-Defined Networking: A Comprehensive Survey. Proceedings of the IEEE, 103(1), 14–76.
 12. Mininet. <http://mininet.org>
 13. Montes de Oca, F.; Galarza, B.; Morales, M.; Encinas, D. “Redes Definidas por Software en Entorno de Cloud Computing”. 6° Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNaIISI 2018). Mar del Plata, Argentina.
 14. Zaccardi, G.; Galarza, B.; Morales, M.; Encinas, D. “Despliegue y ejecución de un cloud privado”. 4° Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNaIISI 2016). Ciudad de Salta, Argentina.
 15. Armanno, M.; Navarro, G.; Morales, M.; Encinas, D. “Utilización de servicios de Cloud Computing y sensores”. 8° Congreso Nacional de Ingeniería Informática/Sistemas de Información (CoNaIISI 2020). San Francisco, Córdoba.
 16. González, N.; Lescano, N.; Pinto, C.; Morales, M.; Encinas, D. “Análisis de rendimiento de IaaS y PaaS”. 8° Congreso Nacional de Ingeniería Informática/Sistemas de Información (CoNaIISI 2020). San Francisco, Córdoba.
 17. Lustre Manual. https://doc.lustre.org/lustre_manual.xhtml#idm140436306123424. Febrero 2021.
 18. Heichler, Jan. "An introduction to BeeGFS." (2014).
 19. Benquerença Mendes, N.; Bond, R.; Morales, M.; Encinas, D. “Rendimiento de sistema de archivos en arquitecturas distribuidas y paralelas”. 8° Congreso Nacional de Ingeniería Informática/Sistemas de Información (CoNaIISI 2020). San Francisco, Córdoba.
 20. E. Párraga, B. León, R. Bond, D. Encinas, A. Bezerra, S. Mendez, D. Rexachs, E. Luque. (2021) Analyzing the I/O Patterns of Deep Learning Applications. Cloud Computing, Big Data & Emerging Topics. JCC-BD&ET 2021. Communications in Computer and Information Science, vol 1444. Springer, Cham. https://doi.org/10.1007/978-3-030-84825-5_1

Software de Base, Modelos y Aplicaciones en Arquitecturas Multiprocesador

Armando De Giusti ⁽¹⁾⁽²⁾ , Marcelo Naiouf⁽¹⁾ , Fernando G. Tinetti ⁽¹⁾⁽³⁾ , Horacio Villagarcía ⁽¹⁾⁽³⁾ , Franco Chichizola⁽¹⁾ , Laura De Giusti⁽¹⁾⁽³⁾ , Enzo Rucci⁽¹⁾⁽³⁾ , Adrián Pousa⁽¹⁾ , Victoria Sanz ⁽¹⁾⁽³⁾ , Diego Montezanti ⁽¹⁾ , Diego Encinas ⁽¹⁾ , Ismael Rodríguez⁽¹⁾ , Sebastián Rodríguez Eguren⁽¹⁾ , Erica Montes de Oca⁽¹⁾ , Juan Manuel Paniego⁽¹⁾ , Martín Pi Puig⁽¹⁾ , César Estrebow⁽¹⁾ , Leandro Libutti⁽¹⁾ , Manuel Costanzo ⁽¹⁾, Joaquín De Antueno⁽¹⁾, Julieta Lanciotti⁽¹⁾, Javier Balladini⁽⁴⁾ 

¹Instituto de Investigación en Informática LIDI (III-LIDI),
Facultad de Informática, Universidad Nacional de La Plata – Comisión de Investigaciones Científicas de la
Provincia de Buenos Aires

²CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas

³CICPBA – Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

⁴Universidad Nacional del Comahue

{degiusti,mnaiouf,fernando,hvw,francoch,ldgiusti,erucci,apousa,vsanz,dmontezanti,dencinas,ismael,seguren,emont
esdeoca,jmpaniego,mpipuig,cesarest,llibutti, mcostanzo,jdeantueno,jlanciotti}@lidi.info.unlp.edu.ar;
javier.balladini@gmail.com

Resumen

El eje de esta línea de I/D lo constituye el estudio de las arquitecturas multiprocesador que integran sistemas distribuidos y paralelos. Incluye como temas centrales:

- Arquitecturas many-core (GPU, procesadores MIC, TPUs), FPGAs, híbridas (diferentes combinaciones de multicores y aceleradores), y asimétricas.
- Desarrollo y evaluación de algoritmos paralelos sobre nuevas arquitecturas y su evaluación de rendimiento computacional y energético.
- Estudio y optimización de código heredado.
- Desarrollo y evaluación de estrategias de resiliencia.
- Modelado y simulación de E/S en HPC.

Palabras clave: *Sistemas Paralelos. Clusters. Arquitecturas asimétricas. GPU, MIC, FPGA, TPU. Eficiencia energética. Resiliencia. Código heredado. E/S paralela.*

Contexto

Se presenta una línea de Investigación que es parte del proyecto “Computación de Alto Desempeño: Arquitecturas, Algoritmos, Métricas de Rendimiento y Aplicaciones en HPC, Big Data, Robótica, Señales y Tiempo Real.” del III-LIDI y de proyectos específicos

apoyados por organismos nacionales e internacionales. También del proyecto “Procesamiento Eficiente de Grandes Datos usando Cómputo de Altas Prestaciones, Edge y Fog” financiado por la Facultad de Informática de la UNLP.

En los temas hay cooperación con varias Universidades de Argentina y se está trabajando con Universidades de América Latina y Europa en proyectos financiados por CyTED, AECID y la OEI (Organización de Estados Iberoamericanos).

Por otra parte, se cuenta con financiamiento de Telefónica de Argentina en Becas de grado y posgrado y se ha tenido el apoyo de diferentes empresas (IBM, Microsoft, Telecom, Intel) en la temática de Cloud Computing.

Se participa en iniciativas como el Programa IberoTIC de intercambio de Profesores y Alumnos de Doctorado en el área de Informática.

Asimismo, el III-LIDI forma parte del Sistema Nacional de Cómputo de Alto Desempeño (SNCAD) del Ministerio de Educación, Cultura, Ciencia y Tecnología de la Nación.

Introducción

Una de las áreas de creciente interés lo constituye el cómputo de altas prestaciones, en el cual el rendimiento está relacionado con dos

aspectos: por un lado, las arquitecturas de soporte, y por otro, los algoritmos que hacen uso de estas.

A la aparición de arquitecturas *many-core* (como las GPU o los procesadores MIC), se ha sumado el uso de FPGAs debido a su potencia de cómputo y rendimiento energético. Su combinación en sistemas HPC da lugar a plataformas híbridas con diferentes características [22].

Lógicamente, esto trae aparejado una revisión de los conceptos del diseño de algoritmos paralelos (incluyendo los mismos lenguajes de programación y el software de base), así como la evaluación de las soluciones que éstos implementan. También resulta necesario investigar las estrategias de distribución de datos y de procesos a fin de optimizar la performance.

Además, el estudio del consumo y la eficiencia energética de los nuevos sistemas paralelos se vuelve tan importante como el de las métricas clásicas (speedup, eficiencia, escalabilidad) debido a los costos económicos y a los problemas operativos asociados [9].

GPUs y Cluster de GPUs

Las GPUs son el tipo de acelerador dominante en la comunidad de HPC hoy en día por su alto rendimiento y bajo costo de adquisición. En la actualidad, tanto NVIDIA como AMD trabajan especialmente en mejorar la eficiencia energética de sus placas y disminuir el alto costo de programación.

La combinación de GPUs con otras plataformas paralelas como clusters y multicores, brindan un vasto conjunto de posibilidades de investigación en arquitecturas híbridas, a partir de diferentes combinaciones como son:

- Máquinas multicore con más de una GPU, que combinan herramientas de programación paralela como OpenMP/CUDA o Pthread/CUDA.
- Cluster de máquinas multicore cada una con una o más placas de GPU, lo que permite combinar OpenMP/MPI/CUDA o Pthread/MPI/CUDA.

Los desafíos que se plantean son múltiples, sobre todo en lo referido a distribución de

datos y procesos en tales arquitecturas híbridas a fin de optimizar el rendimiento de las soluciones.

MIC

En forma reciente Intel brinda una alternativa a partir de la arquitectura MIC (*Many Integrated Core Architecture*). Esta arquitectura permite utilizar métodos y herramientas estándar de programación HPC, lo que los distingue especialmente de las GPUs. De esta forma, se remueven barreras de entrenamiento y se permite focalizar en el problema más que en la ingeniería del software. Xeon Phi es el nombre elegido por Intel para su serie de procesadores many-core. Recientemente, Intel ha lanzado Knights Landing (KNL), la segunda generación de Xeon Phi. A diferencia de sus predecesores que operaban como co-procesadores a través del puerto PCI, los procesadores KNL pueden operar en forma autónoma. Además, integran las nuevas extensiones vectoriales AVX-512 y tecnología de memoria 3D, entre otras características avanzadas [21].

FPGAs

Una FPGA (*Field Programmable Gate Array*) es una clase de acelerador basado en circuitos integrados reconfigurables. La capacidad de adaptar sus instrucciones de acuerdo con la aplicación objetivo le permite incrementar la productividad de un sistema y mejorar el rendimiento energético para ciertos tipos de aplicaciones. Tradicionalmente han sido utilizadas para el procesamiento digital de señales. Sin embargo, en los últimos años, existen dos tendencias claras para extender su uso a otros dominios. En primer lugar, el establecimiento de alianzas estratégicas entre fabricantes de procesadores y de FPGAs para integrar estos dispositivos en arquitecturas híbridas (Intel con Altera; IBM con Xilinx) [10][11]. En segundo lugar, el desarrollo de nuevas herramientas de programación para FPGAs empleando estándares familiares para HPC, con las cuales se espera reducir los tradicionales tiempos y costos de programación [27][31]. Por último, la incorporación de FPGAs a los servicios de

Cloud abre nuevas oportunidades para la explotación de esta clase de aceleradores.

TPUs

Las unidades de procesamiento tensorial (TPU) son una clase de Circuitos Integrados de Aplicación Específica (ASIC) desarrolladas por Google con el propósito de acelerar las cargas de trabajo de aprendizaje automático que requieren las aplicaciones desarrolladas en su framework TensorFlow [29]. Su uso provee una alternativa a otras arquitecturas ya conocidas como CPUs, GPUs y MICs. En ese sentido, interesa analizar las tasas de aceleración y eficiencia energética provistas por esta nueva arquitectura, en comparación con el resto.

Eficiencia energética

La mejora de la eficiencia energética es una de las principales preocupaciones en la informática actual, principalmente a partir de las plataformas con gran cantidad de procesadores. Muchos esfuerzos están orientados a tratar la eficiencia energética y a las metodologías para medirla como ejes de I/D, como una métrica de evaluación relevante.

Entre los puntos de interés pueden mencionarse:

- Análisis de metodologías y herramientas para medir y optimizar el consumo energético.
- Estudio de técnicas para reducir el consumo energético en aplicaciones de HPC de acuerdo con las arquitecturas utilizadas.
- Evaluación de eficiencia energética de diferentes algoritmos y plataformas paralelas.
- Optimización de la eficiencia energética. A partir de los valores de energía que brindan los contadores hardware es posible definir estrategias de programación que lleven a reducir el consumo, manteniendo a su vez el rendimiento en valores aceptables [25].

Código heredado

La mayoría de los programas de simulación numérica que se emplean hoy en día fueron desarrolladas cuando las arquitecturas

paralelas no existían. Es por ello que este conjunto de aplicaciones presenta la oportunidad de desarrollar técnicas y herramientas que permitan optimizar el código, tanto desde el punto de vista computacional como desde la ingeniería de software [28].

Resiliencia

En la actualidad, lograr sistemas resilientes resulta un verdadero desafío considerando el creciente número de componentes, la cercanía a los límites físicos en las tecnologías de fabricación y la complejidad incremental del software. La corrección de las aplicaciones y la eficiencia en su ejecución se torna más importante en HPC debido a los extensos tiempos de ejecución. En ese sentido, resulta relevante desarrollar estrategias de detección y recuperación de fallos, especialmente a través de librerías de software.

Entrada/Salida paralela

A pesar de los avances tecnológicos, las operaciones de E/S en los centros de supercómputo siguen siendo un cuello de botella para determinadas aplicaciones HPC. El rendimiento de un sistema depende de la carga de trabajo (patrones de E/S de las aplicaciones) y de su configuración (hardware y software) [19]. Contar con herramientas que permitan modelar y predecir el comportamiento de este tipo de aplicaciones en HPC resulta fundamental para mejorar su rendimiento [5].

Analizar y diseñar modelos de simulación basados en la arquitectura de E/S paralela, permite disminuir la complejidad y cubrir las exigencias de las aplicaciones en HPC, al poder identificar y evaluar los factores que influyen en las prestaciones [6].

Dispositivos de bajo costo con capacidades para cómputo paralelo

En la actualidad se comercializan placas de bajo costo como Raspberry PI [20] u Odroid [ODR16] que poseen múltiples núcleos simples. Asimismo, existen diversos dispositivos móviles con capacidades similares. Es de interés estudiar cómo explotar

el paralelismo en estos dispositivos para mejorar el rendimiento y/o consumo energético de las aplicaciones [32].

Líneas de Investigación, Desarrollo e Innovación

- Arquitecturas many-core (procesadores MIC, GPU y TPU) y FPGA. Análisis de este tipo de máquinas y de técnicas para desarrollar código optimizado.

- Arquitecturas híbridas (diferentes combinaciones de clusters, multicores, manycores y FPGAs). Diseño de algoritmos paralelos sobre las mismas. Técnicas de resiliencia.

- Exploración de nuevos lenguajes y modelos de programación para HPC.

- Consumo energético en las diferentes arquitecturas de alto desempeño, en particular en relación con los algoritmos paralelos y la configuración de la arquitectura. Análisis de metodologías y herramientas de medición. Modelado y estimación del consumo de potencia de arquitecturas HPC.

- Análisis y desarrollo de modelos e implementación de simuladores de la pila de software de E/S en HPC.

Resultados y Objetivos

Investigación experimental por realizar

- Desarrollar y evaluar algoritmos paralelos sobre nuevas arquitecturas paralelas. Analizar rendimiento, eficiencia energética y costo de programación.

- Analizar las capacidades de lenguajes no convencionales para procesamiento paralelo, considerando rendimiento y costo de programación.

- Realizar el desarrollo de nuevos planificadores de tareas para multicores asimétricos sobre diferentes sistemas operativos con el objetivo de maximizar el rendimiento y minimizar el consumo de energía [24][25].

- Analizar proceso de migración de algoritmos entre arquitecturas diferentes, considerando portabilidad, rendimiento y esfuerzo de programación.

- Calibrar y sintonizar el modelado y simulación de E/S en HPC conseguido para distintos escenarios e infraestructuras

- Desarrollar técnicas de tolerancia a fallas que permitan aumentar la resiliencia de sistemas paralelos y distribuidos.

Resultados obtenidos

- Se compararon soluciones paralelas para el método de cifrado AES en diversas arquitecturas de memoria compartida [26].

- Se evaluaron capacidades de paralelismo de lenguajes no convencionales en arquitecturas multicore, como Python [14] y Rust [3].

- Se realizó un primer estudio de migración de códigos CUDA a DPC++ usando el ecosistema oneAPI [2]

- Se realizaron y analizaron modificaciones al framework TensorFlow para permitir la maleabilidad de hilos [12].

- Se diseñó y desarrolló un prototipo de simulador de transmisión de enfermedades intra-hospitalarias [13].

- Se exploró el uso de diferentes microcontroladores para aplicaciones de aprendizaje automático [7] [8].

- Se evaluó el impacto de las comunicaciones en un cluster heterogéneo de placas RPi [30]

- Se desarrolló y validó un modelo estadístico para consumo de potencia en placas RPi de diferentes generaciones [18].

- Se diseñó la herramienta SEDAR y evaluó su desempeño para detección y recuperación de fallos transitorios [15].

- Se han desarrollado técnicas de modelado y simulación de E/S en HPC que permiten predecir cómo los cambios realizados en los diferentes componentes de éste afectan a la funcionalidad y al rendimiento del sistema [5][6].

- Se analizó el impacto de la E/S en aplicaciones de aprendizaje automático profundo [19].

- Se desarrolló un modelo que permite predecir el consumo energético de un sistema ante diferentes estrategias aplicadas para reducirlo cuando ocurre una falla [16].

- Se realizó un análisis comparativo de rendimiento y eficiencia energética entre las

arquitecturas Intel Xeon Phi KNL vs NVIDIA Pascal usando como caso de estudio el problema de caminos mínimos en un grafo [1].

Organización de Eventos

En el año 2021 se han organizado las IX Jornadas de Cloud Computing, Big Data & Emerging Topics (JCC-BD&ET 2021) en Argentina, con participación de especialistas académicos del país y del exterior y de empresas con experiencia en Cloud Computing [17][4]. En junio de 2022 se organizarán las X JCC-BD&ET.

Formación de Recursos Humanos

Dentro de la temática de la línea de I/D el último año se concluyó 1 Tesis Doctoral, 1 tesis de Maestría y 1 Trabajo Final de Especialización. Al mismo tiempo se encuentran en curso 3 tesis de Doctorado en Ciencias Informáticas y 3 tesis de Maestría.

Además, se participa en el dictado de las carreras de Doctorado en Ciencias Informáticas, y Magíster y Especialización en Cómputo de Altas Prestaciones de la Facultad de Informática de la UNLP (acreditadas por la CONEAU con categoría A, B y A, respectivamente), por lo que potencialmente pueden generarse nuevas Tesis de Doctorado y Maestría, además de Trabajos Finales de Especialización.

Existe cooperación con grupos de otras Universidades del país y del exterior, y hay tesis de diferentes Universidades realizando su Tesis con el equipo del proyecto.

Respecto a las carreras de grado, se dictan por parte de integrantes de la línea de investigación tres materias directamente relacionadas con los temas de ésta: “Taller de Programación sobre GPUs”, “Cloud Computing y Cloud Robotics” y “Conceptos y Aplicaciones en Big Data”.

Referencias

[1] M. Costanzo, E. Rucci, U. Costi, F. Chichizola, and M. Naiouf, “Comparison of HPC Architectures for Computing All-Pairs Shortest Paths. Intel Xeon Phi KNL vs NVIDIA Pascal”. En: Computer Science – CACIC 2020. Revised Selected Papers., Springer

International Publishing, págs. 37-49, doi. 10.1007/978-3-030-75836-3_3, 2021.

[2] M. Costanzo, E. Rucci, C. García-Sánchez, and M. Naiouf. “Early Experiences Migrating CUDA codes to oneAPI”, Short papers of the 9th Conference on Cloud Computing, Big Data & Emerging Topics (JCC-BD&ET 2021), ISBN: 978-950-34-2016-4, págs. 14-18, 2021.

[3] M. Costanzo, E. Rucci, M. Naiouf, and A. D. Giusti, “Performance vs Programming Effort between Rust and C on Multicore Architectures: Case Study in N-Body”, Proceedings of 2021 XLVII Latin American Computing Conference (CLEI), ISBN: 978-1-66549-503-5, págs. 1-10, doi. 10.1109/CLEI53233.2021.9640225, 2021.

[4] A. E. De Giusti, M. Naiouf, L. C. De Giusti, E. Rucci, and F. Chichizola, “Short papers of the 9th Conference on Cloud Computing, Big Data & Emerging Topics” Facultad de Informática (UNLP), ISBN: 978-950-34-2016-4, 2021.

[5] D. Encinas, M. Naiouf, A. De Giusti, S. Méndez, D. Rexachs del Rosario, and E. Luque, “On the Calibration, Verification and Validation of an Agent-Based Model of the HPC Input/Output System” Proceedings of the The Eleventh International Conference on Advances in System Simulation (SIMUL 2019), ISBN: 978-1-61208-756-6, págs. 14-21, 2019.

[6] D. Encinas, S. Mendez, M. Naiouf, A. De Giusti, D. Rexachs del Rosario, and E. Luque, “An Agent-Based Model for Analyzing the HPC Input/Output System” International journal on advances in systems and measurements vol. 13, num. 3 & 4, págs. 192-202, 2020.

[7] C. A. Estrebou, M. Fleming, M. Saavedra, and F. Adra, “MbedML: A Machine Learning Project for Embedded Systems”, Short papers of the 9th Conference on Cloud Computing, Big Data & Emerging Topics (JCC-BD&ET 2021), ISBN: 978-950-34-2016-4, págs. 25-28, 2021.

[8] C. A. Estrebou, M. Fleming, M. D. Saavedra, and F. Adra, “A Neural Network Framework for Small Microcontrollers”, Actas del XXVII Congreso Argentino de Ciencias de la Computación (CACIC 2021), ISBN: 978-987-633-574-4, págs. 51-60, 2021.

[9] W. Feng, X. Feng and R. Ge, "Green Supercomputing Comes of Age," in IT Professional, vol. 10, no. 1, pp. 17-23, Jan.-Feb. 2008, doi: 10.1109/MITP.2008.8.

[10] IBM. “IBM and Xilinx Announce Strategic Collaboration to Accelerate Data Center Applications”. Disponible en <https://www-03.ibm.com/press/us/en/pressrelease/48074.wss>

[11] Intel. “Intel Acquisition of Altera”. Disponible en <http://intelacquiresaltera.transactionannouncement.com>

[12] L. Libutti, L. De Giusti, and M. Naiouf. “PLANIFICACIÓN Y ACELERACIÓN DE ALGORITMOS DE MACHINE LEARNING”.

- Investigación joven (ISSN 2314-3991), vol. 7, num. 2, págs. 552-553, abril de 2021.
- [13] L. Maccallini, D. O. Encinas, and F. Romero. "An Approach to the Modeling and Simulation of Intra-Hospital Diseases". *Journal of computer science and technology* (ISSN 1666-6038), vol. 21, num. 2, págs. 157-169, doi. 10.24215/16666038.21.e14, 2021.
- [14] A. Milla and E. Rucci, "Acelerando código científico en Python usando Numba", *Actas del XXVII Congreso Argentino de Ciencias de la Computación (CACIC 2021)*, ISBN: 978-987-633-574-4, págs. 72-82, 2021.
- [15] D. Montezanti, E. Rucci, A. D. De Giusti, M. Naiouf, D. Rexachs, and E. Luque, "Soft errors detection and automatic recovery based on replication combined with different levels of checkpointing". *Future generation computer systems* (ISSN 0167-739X), vol. 113, págs. 240-254, doi. <https://doi.org/10.1016/j.future.2020.07.003>, 2020.
- [16] M. Morán, J. Balladini, D. Rexachs, and E. Rucci. "Towards Management of Energy Consumption in HPC Systems with Fault Tolerance". *Proceedings from the 2020 IEEE Congreso Bienal de Argentina (ARGENCON)*, págs. 1-8, doi. 10.1109/ARGENCON49523.2020.9505498, 2021.
- [17] Cloud Computing, Big Data & Emerging Topics: 9th Conference, JCC-BD&ET, La Plata, Argentina, June 22-25, 2021, *Proceedings*. Editores: M. Naiouf, E. Rucci, F. Chichizola, and D. Giusti, Springer International Publishing, ISBN: 978-3-030-84824-8, doi. 10.1007/978-3-030-84825-5, 2021.
- [ODR16] Odroid <http://www.hardkernel.com> Accedido 21 de marzo de 2016.
- [18] J. M. Paniego, L. Libutti, M. P. Puig, F. Chichizola, L. De Giusti, M. Naiouf, and A. De Giusti, "Unified Power Modeling Design for Various Raspberry Pi Generations Analyzing Different Statistical Methods". En: *Computer Science – CACIC 2019. communications in Computer and Information Science.*, ISBN: 978-3-030-48325-8, Springer International Publishing, págs. 53-65, 2020.
- [19] E. Párraga, B. León, R. Bond, D. Encinas, A. Bezerra, S. Mendez, D. Rexachs, and E. Luque. "Analyzing the I/O Patterns of Deep Learning Applications". *Proceedings from the 9th Conference on Cloud Computing, Big Data & Emerging Topics (JCC-BD&ET 2021)*, vol. 1444, págs. 3-16, doi. 10.1007/978-3-030-84825-5_1, 2021.
- [20] Raspberry PI. <https://www.raspberrypi.org/>
- [21] Reinders, J., Jeffers, J., Sodani, A. "Intel Xeon Phi Processor High Performance Programming Knights Landing Edition". Morgan Kaufmann Publishers Inc., Boston, MA, USA, 2016
- [22] Rucci, Enzo: "Evaluación de rendimiento y eficiencia energética en sistemas heterogéneos para bioinformática". Tesis de Doctorado en Ciencias Informáticas (Facultad de Informática – UNLP). 2016.
- [23] E. Rucci, M. Naiouf, F. Chichizola, and L. De Giusti "Cloud Computing, Big Data & Emerging Topics. 8th Conference, JCC-BD&ET 2020, La Plata, Argentina, September 8-10, 2020, *Proceedings*". Springer CCIS, ISBN: 978-3-030-61218-4, 2020.
- [24] Juan Carlos Saez, Adrian Pousa, Daniel Chaver, Fernando Castro, Manuel Prieto Matias: "ACFS: A Completely Fair Scheduler for Asymmetric Single-ISA Multicore Systems". In: *ACM SAC 2015 (The 30TH ACM/SIGAPP Symposium on applied computing)*. 2015.
- [25] Saez, J.C., Pousa, A., Rodríguez-Rodríguez, R., Castro, F., Prieto-Matias, M. "PMCTrack: Delivering performance monitoring counter support to the OS scheduler". *The computer journal* Volume 60, Issue 1 January 2017.
- [26] V. Sanz, A. Pousa, M. Naiouf, and A. De Giusti. "Comparison of Hardware and Software Implementations of AES on Shared-Memory Architectures". *Proceedings from the 9th Conference on Cloud Computing, Big Data & Emerging Topics (JCC-BD&ET 2021)*, vol. 1444, págs. 60-70, doi. 10.1007/978-3-030-84825-5_5, 2021.
- [27] Sean Settle: "High-performance Dynamic Programming on FPGAs with OpenCL". In: *IEEE High Performance Extreme Computing Conference*. 2013.
- [28] F. G. Tinetti, M. J. Perez, A. Fraidenraich, and A. E. Altenberg, "Legacy code and parallel computing: updating and parallelizing a numerical model". *The journal of supercomputing* (ISSN 1573-0484), doi. 10.1007/s11227-020-03172-7, 2020.
- [29] Google Inc. "Cloud Tensor Processing Unit (TPU)" Disponible en <https://cloud.google.com/tpu/docs/tpus?hl=es-419>
- [30] O. C. Valderrama Riveros and F. G. Tinetti, "MPI Communication Performance in a Heterogeneous Environment with Raspberry Pi", *Advances in Parallel & Distributed Processing, and Applications. Transactions on Computational Science and Computational Intelligence.*, ISBN: 978-3-030-69984-0, págs. 451-460, doi. 10.1007/978-3-030-69984-0_33, 2021.
- [31] Xilinx Inc. "SDAccel Development Environment". [Online]. Disponible en <http://www.xilinx.com/products/design-tools/software-zone/sdaccel.html>
- [32] P. S. Rodríguez Eguren, F. Chichizola, and E. Rucci, "Análisis del uso de un cluster de Raspberry Pi para cómputo de alto rendimiento". *Actas del XXIV Congreso Argentino de Ciencias de la Computación (CACIC 2018)*, ISBN: 978-950-658-472-6, págs. 134-144, 2018.

Integración de Arquitecturas Edge-Fog-Cloud en Procesamiento Distribuido. Aspectos de Eficiencia y Resiliencia

Armando De Giusti ⁽¹⁾⁽²⁾ , Marcelo Naiouf⁽¹⁾ , Santiago Medina, Joaquín De Antueno⁽¹⁾, Laura De Giusti⁽¹⁾⁽³⁾ , Julieta Lanciotti⁽¹⁾, Fernando G. Tinetti ⁽¹⁾⁽³⁾ , Franco Chichizola⁽¹⁾ , Enzo Rucci⁽¹⁾⁽³⁾ , Adrián Pousa⁽¹⁾ , Victoria Sanz ⁽¹⁾⁽³⁾ , Diego Montezanti ⁽¹⁾ , Diego Encinas ⁽¹⁾ , Ismael Rodríguez⁽¹⁾ , Sebastián Rodríguez Eguren⁽¹⁾ , Juan Manuel Paniego⁽¹⁾ , Martín Pi Puig⁽¹⁾ , Leandro Libutti⁽¹⁾ , Manuel Costanzo ⁽¹⁾

¹Instituto de Investigación en Informática LIDI (III-LIDI),
Facultad de Informática, Universidad Nacional de La Plata – Comisión de Investigaciones Científicas de la
Provincia de Buenos Aires

²CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas

³CICPBA – Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

{degiusti, mnaouf, fernando, smedina, jdeantueno, ldgiusti, jlanciotti, fernando, francoch, erucci, apousa, vsanz, dmontezanti, dencinas, ismael, seguren, jmpaniego, mpipuig, llibutti, mcostanzo,}@lidi.info.unlp.edu.ar

Resumen

El eje de esta línea de I/D lo constituye el estudio de la integración de arquitecturas distribuidas que van desde el nivel de los nodos sensores basados en microcontroladores (Edge Computing), pasando por una capa intermedia de preprocesamiento (Fog Computing) y finalmente la capa de procesamiento en la nube (Cloud Computing).

Los temas centrales son:

- Distribución equilibrada del procesamiento y almacenamiento de datos en cada nivel.
- Análisis de integridad y performance en las comunicaciones, según el grado de distribución del procesamiento.
- Migración de “inteligencia” al nivel “Edge” para reducir consumo y comunicaciones.
- Control de tolerancia a fallos de la arquitectura.
- Evaluación e integración de plataformas y servicios
- Desarrollo y evaluación de aplicaciones que integran niveles de procesamiento.
- Análisis de eficiencia en tiempo, consumo energético y comunicaciones.

Palabras clave: *Sistemas Distribuidos. Cloud Computing. Fog Computing. Edge Computing. IoT, Algoritmos distribuidos. Eficiencia.*

Contexto

Se presenta una línea de Investigación que es parte del proyecto “Computación de Alto Desempeño: Arquitecturas, Algoritmos, Métricas de Rendimiento y Aplicaciones en HPC, Big Data, Robótica, Señales y Tiempo Real.” del III-LIDI y de proyectos específicos apoyados por organismos nacionales e internacionales. También del proyecto “Procesamiento Eficiente de Grandes Datos mediante Cómputo de Altas Prestaciones, Fog y Edge” financiado por la Facultad de Informática de la UNLP y el proyecto “Unidad Inteligente para Control de Consumo Energético” financiado por la Secretaría de Políticas Universitarias y la UNLP.

En el tema hay cooperación con varias Universidades de Argentina y se está trabajando con Universidades de América Latina y Europa en proyectos financiados por ERASMUS, CyTED y la OEI (Organización de Estados Iberoamericanos). En particular con el proyecto “Computación de Altas Prestaciones Eficiente y Segura para Aplicaciones de Servicios de Salud Inteligentes” de la Universidad Autónoma de Barcelona.

Por otra parte, se tiene financiamiento de diferentes empresas de Argentina, en particular en la formación de recursos humanos en la temática de Cloud/Fog y Edge Computing.

Se participa en iniciativas como el Consorcio en temas de Cloud Computing-Big Data y Temas Emergentes, con Universidades de Argentina y España.

Asimismo, el III-LIDI forma parte del Sistema Nacional de Cómputo de Alto Desempeño (SNCAD) del Ministerio de Educación, Cultura, Ciencia y Tecnología de la Nación.

Introducción

El constante aumento de la cantidad de datos a procesar que generan los múltiples sensores “inteligentes” en la actualidad y la necesidad de tiempos de respuesta menores trae como consecuencia plantear nuevas arquitecturas que gestionen esa información.

La integración de capas de procesamiento y servicios, en lo que se denomina Edge y/o Fog Computing ha dado lugar a un nuevo modelo de arquitectura denominado “Edge-Fog Cloud Computing” [1] que trata de generar una serie de ventajas:

- Reducir el tráfico de comunicaciones, con un mayor procesamiento en los mismos nodos sensores o en una capa intermedia, anterior al Cloud.
- Reducir los tiempos de respuesta evitando la comunicación con el Cloud resolviendo en instancias anteriores.
- Resiliencia a fallos de comunicación o cortes de conexión.
- Capacidad de adaptar las instancias de procesamiento en función del contexto, incluyendo mayor procesamiento en los nodos sensores.

Este modelo de arquitectura trae nuevos desafíos, tales como [2], [3], [4], [5]:

- Administración de varios niveles de procesamiento y almacenamiento de datos heterogéneos que aseguren integridad al sistema.
- Definición de protocolos y mecanismos de interoperabilidad entre las diferentes capas de la arquitectura y los servicios en el Cloud.
- Desarrollo de estrategias para el control de fallos.

- Análisis de la distribución óptima de tareas en cada nivel, en función de mejorar los tiempos de respuesta.
- Análisis de estrategias para disminuir el consumo energético.
- Estudio e integración de diferentes plataformas, aplicaciones y servicios.
- Seguridad e integridad en los datos.
- Estudios de escalabilidad (no lineal) con el crecimiento de los sensores en la capa “Edge”.

Cloud Computing

Cloud Computing, proporciona grandes conjuntos de recursos físicos y lógicos (como pueden ser infraestructura, plataformas de desarrollo, almacenamiento y/o aplicaciones), fácilmente accesibles y utilizables por medio de una interfaz de administración web, con un modelo de arquitectura “*virtualizada*” [6] [7]. Estos recursos son proporcionados como servicios (“*as a service*”) y pueden ser dinámicamente reconfigurados para adaptarse a una carga de trabajo variable (escalabilidad), logrando una mejor utilización y evitando el sobre o sub dimensionamiento (elasticidad) [8].

En nuestro modelo de procesamiento distribuido que integra desde sensores a la nube, Cloud Computing se reserva para el procesamiento centralizado de algoritmos complejos, con gran volumen de datos. Parte de estos datos podrán haber sido “preprocesados” en las capas “Edge” o “Fog”. [9] [10] [11] [12].

Fog Computing

El modelo de Fog Computing surge como respuesta al crecimiento de los desarrollos relacionados con el Internet de las Cosas. Estos desarrollos requieren procesamiento en la nube, pero presentan características que hacen complejo su armado haciendo uso exclusivo de tecnologías de Cloud Computing. Fog Computing consiste entonces en una plataforma intermedia que provee procesamiento, almacenamiento y servicios de comunicación en red entre los dispositivos “Edge” que adquieren los datos en el modelo IoT. Las características de Fog Computing son

las de una capa intermedia orientada a [13] [14] [15]:

- Aplicaciones de tiempo real distribuidas.
- Gran número de nodos, que puede escalar dinámicamente.
- Heterogeneidad de los nodos.
- Movilidad.
- Red predominante de comunicaciones inalámbricas.
- Baja latencia y conocimiento de la ubicación.

Edge Computing

El modelo de “Edge” computing está impuesto por el crecimiento exponencial del número de dispositivos sensores con inteligencia local disponibles. Internet “de las cosas” (IoT) crece y hoy hablamos de 50.000 millones de dispositivos conectados a Internet, con un tráfico del orden de 800 Zbytes. Se requiere entonces capacidad de procesamiento cerca de los sensores, integración de datos locales y geográficamente ubicados y posibilidad de pre procesar los mismos y enviarlos con un menor overhead de comunicaciones a las capa superiores (Fog o Cloud). [16] [17] [18] [19]. Las ventajas del modelo Edge son directas [20] [21] :

- Respuestas automáticas al usuario, en tiempo real, utilizando su capacidad local.
- Disminución del tráfico de datos.
- Mayor seguridad por el procesamiento local.
- Disminución del consumo energético, por las características de los componentes.
- Mejora de la eficiencia global para un sistema distribuido, débilmente acoplado.

Aplicaciones de tiempo real.

Este nuevo modelo Edge-Fog Cloud Computing es especialmente aplicable a problemas de tiempo real (Cloud robotics, Vehículos autónomos, Monitoreos de salud personalizados, etc.), al permitir que las capas cercanas a los sensores y al usuario resuelvan en menor tiempo y con menor overhead de

comunicaciones, la respuesta “inmediata”, dejando para el Cloud el procesamiento de los datos masivos “off-line” (por ejemplo para ajustar/perfeccionar un modelo de comportamiento que luego se transforma en ajustes a los algoritmos en las capas Edge y Fog. [22] [23].

Eficiencia energética

La mejora de la eficiencia energética es un tema central en la informática actual, principalmente a partir de las plataformas con gran cantidad de procesadores. Muchos esfuerzos están orientados a tratar la eficiencia energética como eje de I/D, como métrica de evaluación, y también a la necesidad de metodologías para medirla. En el caso del modelo Edge-Fog Cloud Computing, hay una mejora del consumo por dos efectos:

- Normalmente los procesadores de la capa “Edge” son de bajo consumo.
- Al bajar el tráfico de comunicaciones, también se baja el consumo energético asociado al mismo.

Líneas de Investigación, Desarrollo e Innovación

- Arquitecturas aplicables en Edge Computing. Modelos de referencia.
- Arquitecturas aplicables en Fog Computing. Modelos de referencia.
- Administración de recursos y datos en Edge y Fog computing. Integración de los mismos.
- Seguridad e integridad en los datos.
- Vinculación de las capas Edge y Fog con el Cloud. Servicios requeridos.
- Vinculación de las capas Edge y Fog con plataformas dedicadas al Internet de las Cosas.
- Estrategias de distribución óptima de procesamiento entre capas.
- Escalabilidad en aplicaciones Edge-Fog-Cloud.
- Migración de inteligencia al nivel Edge para reducir consumo y comunicaciones.
- Métricas de eficiencia considerando tiempo de respuesta / costo

comunicaciones / consumo energético / resiliencia.

- Desarrollo de diferentes nodos: nodos sensores, robots, drones, etc.
- Algoritmos colaborativos en tiempo real integrando las tres capas.
- Aplicaciones: sistemas inteligentes distribuidos para reducir el consumo energético.

Resultados y Objetivos

Investigación experimental por realizar

- Análisis comparativo de tecnologías y plataformas empleadas en Edge/Fog Computing.
- Análisis comparativo de los servicios para arquitecturas Edge y Fog.
- Estudio de protocolos de comunicación con el Cloud, desde el nivel de Edge y/o Fog.
- Análisis de integridad y performance en comunicaciones de larga y corta distancia, en función del escalado de la arquitectura de sensores.
- Estudios de consumo energético en aplicaciones de Edge/Fog/Cloud.
- Estudio comparativo de la distribución de carga de procesamiento en aplicaciones distribuidas (en particular de tiempo real).
- Integración de aplicaciones móviles y relación entre las capas Edge y Fog.
- Análisis de estrategias para el control de fallos en las diferentes capas de la arquitectura.
- Aplicaciones: robots / drones colaborativos trabajando en comunicación con una capa Fog y con el Cloud.
- Aplicaciones: Sistema de tableros inteligentes para optimización de consumo energético en edificios distribuidos.

Resultados obtenidos

- Se han configurado, desplegado y analizado plataformas orientadas a la gestión de recursos en Edge y Fog Computing.

- Se han realizado comparaciones entre diferentes tipos de servicios para la integración con el Cloud.
- Se han estudiado protocolos y tráfico de comunicaciones en aplicaciones distribuidas en tiempo real que requieren interactuar con el Cloud.
- Se desarrolló un modelo de tablero inteligente para reducción del consumo y su integración con la capa Fog y comunicación con la nube.
- Se está estudiando el tema de consumo distribuido, en el caso de los tableros inteligentes, de los robots distribuidos y del trabajo con drones.

Organización de Eventos

En el año 2021 se han organizado las IX Jornadas de Cloud Computing, Big Data & Emerging Topics (JCC-BD&ET 2021) en Argentina, con participación de especialistas académicos del país y del exterior y de empresas con experiencia en Cloud Computing. En junio de 2022 se organizarán las X JCC-BD&ET.

Formación de Recursos Humanos

Dentro de la temática de la línea de I/D están en curso 3 Tesis de Maestría.

Asimismo se desarrollaron 2 trabajos finales de alumnos.

Además, se participa en el dictado de las carreras de Doctorado en Ciencias Informáticas, y Magíster y Especialización en Cómputo de Altas Prestaciones de la Facultad de Informática de la UNLP (acreditadas por CONEAU), por lo que potencialmente pueden generarse Tesis de Doctorado y Maestría y Trabajos Finales de Especialización.

Existe cooperación con grupos de otras Universidades del país y del exterior con posibilidad de realizar Tesis en colaboración.

Respecto a las carreras de grado, se dictan por parte de integrantes de la línea de investigación dos asignaturas directamente relacionadas con los temas de la misma: "Cloud Computing y Cloud Robotics" y "Conceptos y Aplicaciones en Big Data".

Asimismo, todos los años se desarrollan proyectos con alumnos, relacionados básicamente con aplicaciones de tiempo real con robots y drones.

Referencias

- [1] Mohan N., Kangasharju J. "Edge-Fog Cloud: A Distributed Cloud for Internet of Things Computations", 2016 Cloudification of the Internet of Things (CIoT), Paris, France, 2016, pp. 1-6.
- [2] P. Garcia Lopez et al. "Edge-centric computing: Vision and challenges," *SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 5, pp. 37–42, Sep. 2015.
- [3] M. Yannuzzi et al. "Key ingredients in an iot recipe: Fog computing, cloud computing, and more fog computing," in *IEEE CAMAD*, 2014.
- [4] K. Hong et al. "Mobile fog: A programming model for large-scale applications on the internet of things," in *ACM SIGCOMM Workshop on Mobile Cloud Computing*, 2013.
- [5] A. Chandra, J. Weissman, and B. Heintz, "Decentralized edge clouds," *Internet Computing, IEEE*, vol. 17, no. 5, pp. 70–73, Sept 2013.
- [6] E. Rucci, M. Naiouf, F. Chichizola, and L. De Giusti "Cloud Computing, Big Data & Emerging Topics. 8th Conference, JCC-BD&ET 2020, La Plata, Argentina, September 8-10, 2020, Proceedings". Springer CCIS, ISBN: 978-3-030-61218-4, 2020.
- [7] Xing, Y., Zhan, Y.: "Virtualization and Cloud Computing". In: Proceedings pp.305-312, Springer Link. ISBN 978-3-642-27323-0. (2012). Morgan Kaufmann. 2013.
- [8] Velte, A.T., Velte, T.J., Elsenpeter, R.: "Cloud Computing: A Practical Approach". McGraw Hill Professional. 2009.
- [9] Ashkan Yousefpour, Caleb Fung, Tam Nguyen, Krishna Kadiyala, Fatemeh Jalali, Amirreza Niakanlahiji, Jian Kong, Jason P. Jue. "All one needs to know about fog computing and related edge computing paradigms: A complete survey", *Journal of Systems Architecture*, Volume 98, 2019, Pages 289-330, ISSN 1383-7621, <https://doi.org/10.1016/j.sysarc.2019.02.009>.
- [10] Pi Puig M. et al. (2019) Intelligent Distributed System for Energy Efficient Control. In: Naiouf M., Chichizola F., Rucci E. (eds) *Cloud Computing and Big Data. JCC&BD 2019. Communications in Computer and Information Science*, vol 1050. Springer, Cham
- An Overview on Edge Computing Research
- [11] Keyan Cao, Yefan Liu, Gongjie Meng, Quimeng Sun "An Overview on Edge Computing Research" in *IEEE Access*, vol. 8, pp. 85714-85728, 2020.
- [12] W. S. Shi, X. Z. Zhang, and Y. F. Wang, "Edge computing: State-of-the-art and future directions," *J. Comput. Res. Develop.*, vol. 56, no. 1, pp. 1_21, 2019.
- [13] X. Hong and Y. Wang, "Edge computing technology: Development and countermeasures," *Chin. J. Eng. Sci.*, vol. 20, no. 2, p. 20, 2018.
- [14] D. Evans. "The Internet of Things How The Next Evolution of the Internet is Changing Everything". Available: <https://www.researchgate.net/publication/30612290>
- [15] M. Satyanarayanan, "The emergence of edge computing," *Computer*, vol. 50, no. 1, pp. 30_39, Jan. 2017.
- [16] Y. Q. Gao, H. Bguan, and Z. W. Qi, "Service level agreement based energy-Efficient resource man agreement in cloud data centers," *Comput. Elect. Eng.*, vol. 40, no. 5, pp. 1621_1633, 2014.
- [17] W. Shi, H. Sun, J. Cao, Q. Zhang, and W. Liu, "Edge computing-an emerging computing model for the Internet of everything era," *J. Comput. Res. Develop.*, vol. 54, no. 5, pp. 907_924, May 2017.
- [18] J. de Antueno, S. Medina, L. De Giusti and A. De Giusti, "Analysis, Deployment and Integration of Platforms for Fog Computing", *Journal of Computer Science and Technology*, 20(2), e12, October 2020.
- [19] F. Bonomi, R. Milito, J. Zhu and S. Addepalli, "Fog Computing and Its Role in the Internet of Things", in *MCC '12: Proceedings of the first edition of the MCC workshop on Mobile cloud computing. Association for Computing Machinery, New York, NY, United States*, 2012.
- [20] M. Asemami, F. Jabbari, F. Abdollahei and P. Bellavista, "A Comprehensive Fog-enabled Architecture for IoT Platforms", *High-Performance Computing and Big Data Analysis. TopHPC 2019. Communications in Computer and Information Science, vol 891. Springer, Cham*, 2019.
- [21] W. S. Aung and S. Aung Nyein Oo, "Monitoring and Controlling Device for Smart Greenhouse by using Thingier.io IoT Server". *International Journal of Trend in Scientific Research and Development*, 2019.
- [22] Dr. S. K. Selvaperumal, W. Al-Gumaei, R. Abdulla and V. Thiruchelvam, "Integrated Wireless Monitoring System Using LoRa and Node-Red for University Building". *Journal of Computational and Theoretical Nanoscience, Volume 16, Number 8, American Scientific Publishers*, 2019.
- [23] M. Pi Puig, J. M. Paniego, S. Medina, S. Rodriguez Eguren, L. Libutti, J. Lanciotti, J. de Antueno, C. Estrebou, F. Chichizola and L. De Giusti, "Intelligent Distributed System for Energy Efficient Control", *Cloud Computing and Big Data. JCC&BD 2019. Communications in Computer and Information Science, vol 1050. Springer, Cham*, 2019.

Migración de aplicaciones Monolíticas a entornos distribuidos Serverless

Nelson Rodríguez¹, María Murazzo¹, Diego Medel¹, Daniel Arias Figueroa², Lorena Parra¹, Ana Laura Molina¹, Adriana Martín¹, Hernán Atencio³, Martín Gómez³
Guillermo Casasola³

¹ Departamento e Instituto de Informática - F.C.E.F. y N. - U.N.S.J.

² Consejo de Investigación, FCE, Universidad Nacional De Salta

³ Alumno Avanzado Licenciatura en Sistemas de Información y Cs. de la Computación - F.C.E.F. y N. - U.N.S.J.

Complejo Islas Malvinas. Cereceto y Meglioli. 5400. Rivadavia. San Juan, 0264 4234129

nelson@iinfo.unsj.edu.ar, marite@unsj-cuim.edu.ar, vdiego.unsj@hotmail.com,
daaf@cidia.unsa.edu.ar, lorenaparra152@yahoo.com.ar, almm95@gmail.com, fgsanchez@unsj-
cuim.edu.ar, adrianamartin1@gmail.com, hernan.atencio.98@gmail.com,
martinsj0811@gmail.com, guillecasasola@gmail.com

Resumen

Serverless Computing es una arquitectura o modelo de ejecución en el cloud, alternativo al modelo tradicional. Ofrece numerosas ventajas sobre una arquitectura monolítica, como aportar agilidad, innovación, un mejor escalado automático, flexibilidad en el desarrollo y una mejor evaluación y control de los costos. Surgió como una evolución de microservicios corriendo en contenedores e implementando funciones, por lo cual a veces se lo denomina función como servicio. Las ventajas que presenta este modelo de computación, promueven a que sea conveniente migrar aplicaciones montadas sobre el cloud tradicional a una arquitectura serverless. Estas aplicaciones se deben llevar a una arquitectura de microservicios, donde se ejecutan funciones, cada una de las cuales de forma independiente y conducida por eventos. Si bien existen varias estrategias y propuestas metodológicas para llevar a cabo la migración, surgen numerosos desafíos y problemas a resolver, debido a que el desarrollo es completamente diferente, además se deben aplicar técnicas de observación y monitorear el progreso de la migración, por lo tanto toda la problemática expresada es el motivo de la presente línea de investigación.

Palabras clave: *Serverless Computing, Distributed Computing, Software Architecture, Cloud Computing*

Contexto

El presente trabajo se encuadra dentro del área de I/D Procesamiento Distribuido y Paralelo y es una de las líneas de investigación internas, del proyecto: Computación Serverless para el tratamiento de datos provenientes de dispositivos de IoT, cuya propuesta ha sido aprobada y está en desarrollo para el período 2020-2021, y se ha extendido un año más. Asimismo el grupo de investigadores viene trabajando en proyectos relacionados con la computación distribuida y de alta performance desde hace más de 21 años. Como continuación del proyecto anterior: Orquestación de Servicios para la Continuidad Edge al Cloud, se continúa el trabajo con investigadores de otras universidades, lo cual favorece notablemente a todas las instituciones participantes.

Introducción

Cloud Computing es una arquitectura o modelo de provisión de servicios bastante adoptado y aceptado, sin embargo administrar

estos servicios no es una tarea fácil en absoluto. Los autores de [1] han abordado varios desafíos al administrar un entorno de nube por parte de un usuario, como la disponibilidad, el equilibrio de carga, el escalado automático, la seguridad, la supervisión, etc.

Estos desafíos han llevado a introducir otro modelo informático en el Cloud, que se denomina computación Serverless.

Tiene como objetivo fundamental aliviar varios problemas que tienen los desarrolladores en el Cloud tradicional, permitiéndoles concentrarse solo en la funcionalidad principal de su producto [2].

La computación Serverless es una tecnología con un impacto creciente en nuestra sociedad y una mayor adopción tanto por parte de la academia como de la industria [3]. Es un paradigma en el que las aplicaciones de software se descomponen en múltiples funciones independientes sin estado [4] [5]. Las funciones se ejecutan en contenedores en respuesta a acciones desencadenantes (como interacciones de usuario, eventos de mensajería o cambios en la base de datos), se pueden escalar de forma independiente y pueden ser efímeras.

En este enfoque, casi todas las preocupaciones operativas son abstraídas lejos de los desarrolladores. Los cuales en principio simplemente escriben código e implementan sus funciones en una plataforma sin servidor. La plataforma se encarga de la ejecución de la función, el almacenamiento y la infraestructura de contenedor, redes y tolerancia a fallas. Adicionalmente, también se encarga de escalar las funciones según la demanda real.

En la mayoría de los casos, se puede escribir funciones en el lenguaje favorito del programador (Node.js, Python, Go, Java y más) y utilizar herramientas de contenedor y serverless, como AWS SAM o la CLI de Docker, para compilar, probar e implementar las funciones.

La computación serverless ha sido identificada como un enfoque prometedor para varias aplicaciones, como el análisis de datos en el edge [6]. En consecuencia, una plataforma maneja el ciclo de vida, la ejecución y escalada

de las funciones reales; estas necesitan correr solo cuando son invocadas o activadas. Por lo tanto, el mayor beneficio son las pocas preocupaciones operativas y de gestión y la utilización eficiente de los recursos [7].

Un modelo basado en funciones es particularmente adecuado para ráfagas, uso de CPU intensivo, cargas de trabajo granulares. Actualmente, los casos de uso de FaaS varían ampliamente, incluido el procesamiento de datos, el procesamiento de flujo, la computación de borde (IoT) y la computación científica [8] [9], y es probable que otros casos de uso surjan en un futuro.

El término "sin servidor" es confuso ya que hay hardware y procesos de servidor que se ejecutan, pero la diferencia en comparación con los enfoques tradicionales es que la organización que construye y admite una aplicación "sin servidor" no se ocupa de ese hardware o esos procesos, debido a que están subcontratando esta responsabilidad.

El término comenzó a utilizarse en 2012 [10].

El término se hizo más popular en 2015, luego del lanzamiento de AWS Lambda,

Existen varias definiciones, se ha tomado como referencia la elaborada por el grupo de investigación The SPEC Cloud, que la describe de la siguiente manera:

La computación serverless es una forma de computación en el cloud que permite a los usuarios ejecutar eventos y aplicaciones facturadas de forma granular, sin tener que abordar la lógica operativa [11].

Los desarrolladores se centran en abstracciones de alto nivel (por ejemplo, funciones, consultas y eventos) y en crear aplicaciones que se asignan a recursos concretos y servicios de soporte. Esto permite que los desarrolladores se enfoquen en la lógica empresarial y en las formas de interconectar elementos de la lógica empresarial en flujos de trabajo complejos. Mientras tanto, los proveedores de servicios se aseguran de que las aplicaciones están alojadas en contenedores, desplegadas, aprovisionadas y disponibles bajo demanda, mientras se factura al usuario solo por los recursos utilizados [12].

La computación serverless se puede identificar como resultado de la unión de Cloud y Microservices Architecture. Pero la evolución a Serverless ha pasado por varias etapas y está en permanente crecimiento.

Migración de aplicaciones Monolíticas a Serverless Computing

El término arquitectura monolítica se deriva de la era histórica cuando los edificios antiguos fueron tallados, fundidos o excavados a partir de una sola pieza de material. Por ejemplo, las iglesias monolíticas de Etiopía y Pancha Rathas en la India fueron talladas en una sola pieza de roca. La mayor desventaja era que si se realizaban pequeños trabajos de reparación afectaba a toda la arquitectura.

Del mismo modo, en el contexto de desarrollo de software, los componentes de la arquitectura monolítica están fuertemente acoplados. Por lo tanto, si se planea escalar, eliminar errores o realizar ligeros cambios en cualquier componente, toda la aplicación se verá afectada. Eso significa que, en tales casos, se debe reescribir la aplicación desde cero.

Serverless presenta muchas ventajas sobre la arquitectura monolítica, como son: recuperación más rápida de fallos, reducción del mantenimiento del servidor, reutilización de código y libertad para codificar en diferentes lenguajes. Por lo tanto es conveniente realizar la migración a este entorno.

La migración seguramente implicará una vasta reescritura del código existente, a diferencia de, por ejemplo, si está migrando una aplicación tradicional basada en servidor para que se ejecute dentro de contenedores donde los cambios generalmente se limitan al nivel de infraestructura. Se debe realizar un análisis de costo-beneficio antes de continuar.

Otro aspecto muy importante es la migración de la base de datos, por ejemplo, si se tiene la aplicación en MongoDB, surgen interrogantes sobre cómo realizar la migración, por ejemplo a DynamoDB (si se usa AWS Lambda) y todo lo que ello implica: sincronización de datos, autenticación, tiempo de transición en que ambos están en uso, etc.

En el inicio de la migración, se debe desacoplar el sistema monolítico, dividiéndolo en una serie de servicios bien definidos y poco acoplados [13].

Serverless, son aplicaciones sin estado, sin embargo las aplicaciones monolíticas, por lo general tienen estado. En este caso se deben implementar mecanismos para proveer de estado a la aplicación.

El arranque en frío, es una característica de Serverless, dado que la función a ejecutar debe cargarse en el contenedor y esto lleva una demora que puede afectar la performance, en el caso de que exista algún requerimiento temporal, debe ser tenido en cuenta.

El diseño de aplicaciones distribuidas (en este caso rediseño), requiere que los programadores tomen en cuenta la totalidad de los aspectos operativos incluyendo confiabilidad, mantenimiento y sobreprovisión de recursos, estos deben ser conocidos en profundidad.

Al ser código heredado de una aplicación monolítica, se debe considerar si se va a realizar esta migración de forma gradual, o se va a implementar todo el sistema. En el caso de una migración gradual: Los componentes de la aplicación se reemplazan por versiones serverless de ellos, una a la vez, mientras los usuarios utilizan esta versión híbrida.

Un plan de migración está determinado por: El estado actual de su organización, el estado actual de la aplicación y el estado deseado. Existen tres estrategias generales de migración para crear una aplicación sin servidor:

- Leapfrog
- Organic
- Strangler

La estrategia Leapfrog se salta las fases intermedias y se mueve directamente a una arquitectura de Cloud Serverless.

Con la estrategia Orgánica, se "trasladan" los programas locales al Cloud. Las aplicaciones actuales se mantienen ejecutándose en instancias Amazon EC2 o Amazon ECS (por ejemplo si la migración fuera en AWS).

En la estrategia **strangler**, se descompone los programas monolíticos mediante el establecimiento de API y componentes basados en eventos. Combina la interfaz de usuario y el código de acceso a los datos y

reemplazan lentamente los componentes heredados. Permite un desarrollo más rápido de nuevas características con menos riesgo que la estrategia Leapfrog. Strangler es la estrategia más común [14].

Líneas de Investigación, Desarrollo e Innovación

La investigación sobre esta línea de trabajo ha comenzado hace un año aproximadamente a partir de un análisis documental sobre revisiones sistemáticas [15], surveys [16] [17] [18] y otros trabajos, que permitió profundizar sobre el estado del arte en la migración de aplicaciones monolíticas a Serverless. Esto permitió encontrar cuáles son los problemas científicos, consideraciones y desafíos que se van a tratar de solucionar.

Posteriormente la investigación se conducirá de forma hipotética mixta (experimental y deductiva). Esta forma de trabajo permitirá analizar la estrategia en diferentes tipos de migraciones según el tipo de aplicación, enfrentar problemas y desafíos y construir un conjunto de buenas prácticas. El método a utilizar será empírico analítico, pero puede variar en función de cada problema en particular.

Resultados y Objetivos

Resultados Obtenidos

Durante los últimos catorce años se trabajó en el área de Computación de Altas Prestaciones y distribuidas, en particular sobre análisis de diversas arquitecturas paralelas y distribuidas, tales como: Cloud Computing, Cluster de commodity, arquitecturas distribuidas y paralelas de bajo costo y fog computing. El proyecto marco de esta línea de investigación, se inició hace dos años, el mismo tiene como temática principal a Serverless Computing. A partir de distintos análisis y debates en el grupo, se inició la línea de investigación del presente trabajo. El grupo ha realizado varias publicaciones en esta área: ocho trabajos de investigación en Congresos y Jornadas, se realizaron tres publicaciones en revistas

científicas y se transfirieron los resultados mediante conferencias en eventos científicos. Se han aprobado dos tesinas de grado, se incorporó un becario de investigación categoría alumno y otra beca está en evaluación.

Objetivos

Los objetivos del grupo de investigación en esta línea de conocimiento son los siguientes:

- Realizar aportes desde la academia a la temática de la presente propuesta, debido a que existen pocos trabajos científicos publicados. Por otro lado, si bien la industria ha realizado su aporte (aunque son también pocas publicaciones) son menos rigurosos en sus conclusiones, procedimientos y métodos.
- Desarrollar un conjunto coherente de buenas prácticas, que permitan ofrecer una guía o soporte metodológico para que el equipo de desarrollo lleve adelante sus tareas.
- Profundizar en el análisis y tratamiento de problemas abiertos o aún no resueltos, para lograr una efectiva migración.

Formación de Recursos Humanos

El equipo de trabajo de esta línea de investigación está compuesto de ocho investigadores que figuran en este trabajo de las universidades Nacional de San Juan y Nacional de Salta y dos alumnos de grado. Además, el proyecto marco donde se está desarrollando esta propuesta incluye a tres investigadores más de la Nacional de San Luis, de la Universidad Champagnat y de la Universidad Nacional de San Juan y tres alumnos de grado.

Se está desarrollando una tesis doctoral sobre paralelismo híbrido y Big Data, una tesis de maestría en áreas afines y seis tesinas de grado en el área de Serverless computing, Concurrencia y Computación distribuida, en particular dos sobre migración de aplicaciones monolíticas. Además se espera aumentar el número de publicaciones. Por otro lado también se prevé la divulgación de varios temas investigados por medio de cursos de postgrado y actualización o publicaciones de

divulgación y asesoramiento a empresas y otros organismos del estado.

Referencias

- [1] Jonas E, Schleier-Smith J, Sreekanti V, Tsai C-C, Khandelwal A, Pu Q, Shankar V, Carreira J, Krauth K, Yadwadkar N, Gonzalez JE, Popa RA, Stoica I, Patterson DA (2019) Cloud Programming Simplified: A Berkeley View on Serverless Computing. <http://arxiv.org/abs/1902.03383>. Accessed 6 Jan 2021
- [2] S. Eismann et al., ‘Serverless Applications: Why, When, and How?’ (2021), *IEEE Software*, vol. 38, no. 1, pp. 32–39, Jan. 2021, doi: 10.1109/MS.2020.3023302.
- [3] IDC, FutureScape: Worldwide IT Industry 2019 Predictions," <https://www.idc.com/getdoc.jsp?containerId=US44403818>, (2018).
- [4] Adzic, G., & Chatley, R. (2017). Serverless computing: economic and architectural impact. In *Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering* (pp. 884-889). ACM.
- [5] Baldini, I., Castro, P., Chang, K., Cheng, P., Fink, S., Ishakian, V. & Suter, P. (2017). Serverless computing: Current trends and open problems. In *Research Advances in Cloud Computing* (pp. 1-20). Springer, Singapore.
- [6] Nastic, S., Rausch, T., Scekcic, O., Dustdar, S., Gusev, M., Koteska, B. & Prodan, R. (2017). A serverless real-time data analytics platform for edge computing. *IEEE Internet Computing*, 21(4), 64-71.
- [7] Mohanty, S. K., Premsankar, G., & Di Francesco, M. (2018). An Evaluation of Open Source Serverless Computing Frameworks. In *CloudCom* (pp. 115-120).
- [8] Gottlieb, N. (2016). State of the Serverless Community Survey Results. <https://serverless.com/blog/state-of-serverless-community/>.
- [9] Jonas, E., Pu, Q., Venkataraman, S., Stoica, I., & Recht, B. (2017). Occupy the cloud: Distributed computing for the 99%. In *Proceedings of the 2017 Symposium on Cloud Computing* (pp. 445-451). ACM.
- [10] Fromm, K. (2012). <https://readwrite.com/2012/10/15/why-the-future-of-software-and-apps-is-serverless/>
- [11] Van Eyk, E., Iosup, A., Seif, S., & Thömmes, M. (2017). The SPEC cloud group's research vision on FaaS and serverless architectures. In *Proceedings of the 2nd International Workshop on Serverless Computing* (pp. 1-4). ACM.
- [12] Van Eyk, E., Toader, L., Talluri, S., Versluis, L., Uță, A., Iosup, A. (2018). Serverless is more: From paas to present cloud computing. *IEEE Internet Computing*, 22(5), 8-17.
- [13] Alireza Goli, Omid Hajihassani, Hamzeh Khazaei*, Omid Ardakanian, Moe Rashidi and Tyler Dauphinee. Migrating from Monolithic to Serverless: A FinTech Case Study. In *ACM/SPEC International Conference on Performance Engineering Companion (ICPE '20 Companion)*, April 20–24, 2020, Edmonton, AB, Canada. ACM, New York, NY, USA, 6 pages.
- [14] Mustafa Osama. Migrating to serverless. *Migrating to Serverless*. (2021). Technology Geek (osamaoracle.com)
- [15] Kjorveziroski, V.; Filiposka, S.; Trajkovik, V. IoT Serverless Computing at the Edge: A Systematic Mapping Review. *Computers* (2021), <https://doi.org/10.3390/computers10100130>
- [16] Hassan et al. Survey on serverless computing *Journal of Cloud Computing: Advances, Systems and Applications* (2021) 10:39 <https://doi.org/10.1186/s13677-021-00253-7>
- [17] Zijun Li, Linsong Guo, Jiagan Cheng, and Quan Chen, The Serverless Computing Survey: A Technical Primer for Design Architecture. (2022) *ACM Computers Survey*. <https://doi.org/10.1145/3508360>
- [18] Zhe Li et al. A Survey of Cost Optimization in Serverless Cloud Computing (2021) *Journal Phys.: Conf. Ser.* 1802 032070

HPC Serverless para tratamiento de datos provenientes del IoT

Nelson Rodríguez¹, María Murazzo¹, Diego Medel¹, Fabiana Piccoli², Lorena Parra¹, Ana Laura Molina¹, Adriana Martín¹, Miguel Méndez Garabetti³, Pablo Gómez⁴, Joaquín Lebeti⁴

¹ Departamento e Instituto de Informática - F.C.E.F. y N. - U.N.S.J.

² Departamento de Informática - F.C.F.M. y N – UNSL

³ Instituto de Investigaciones, Facultad de Informática y Diseño, Universidad Champagnat

⁴ Alumno Avanzado Licenciatura en Sistemas de Información y Cs. de la Computación - F.C.E.F. y N. - U.N.S.J.

Complejo Islas Malvinas. Cereceto y Meglioli. 5400. Rivadavia. San Juan, 0264 4234129

nelson@iinfo.unsj.edu.ar, marite@unsj-cuim.edu.ar, vdiego.unsj@hotmail.com, mpiccoli@unsl.edu.ar, lorenaparra152@yahoo.com.ar, almm95@gmail.com, adrianamartin1@gmail.com, mendez-garabettimiguel@uch.edu.ar, pablo.gomez.allende@gmail.com, lebejoaquin@gmail.com

Resumen

Se ha demostrado que la ejecución de aplicaciones de HPC en el cloud es una opción viable a las arquitecturas paralelas o distribuidas convencionales, las cuales requieren un alto grado de administración, así como un pobre escalado de recursos.

El enfoque tradicional para un usuario usualmente es utilizar al proveedor de Cloud para aprovisionar máquinas virtuales (VM) empleándolas de manera similar a una infraestructura local, con el consiguiente problema de la administración de recursos sumado a la degradación de la performance de las aplicaciones por la contextualización de los ambientes virtualizados.

Serverless computing, permite a un usuario ejecutar código escrito en el lenguaje de programación de su elección, sin tener que aprovisionar primero una máquina virtual. Por otro lado, la elasticidad, disponibilidad, escalabilidad y la tolerancia a fallas son proporcionadas de manera transparente por el proveedor cloud. De esta manera es posible disminuir la complejidad de la administración de la infraestructura para el desarrollador, permitiéndole que se centre en la lógica de la

aplicación. Y además surgen ventajas económicas, al pagar solo por el tiempo de uso.

La presente línea de investigación se centra en el desafío de evaluar el costo, no solo monetario sino también de performance, de migrar aplicaciones de HPC a entornos serverless. Esta evaluación permitirá que se pueda tomar la decisión que infraestructura se usará con la finalidad que se obtenga el mejor beneficio de performance.

Palabras clave: *Serverless Computing, HPC, Cloud Computing, IoT*

Contexto

El presente trabajo se encuadra dentro del área de I/D Procesamiento Distribuido y Paralelo y es una de las líneas de investigación internas, del proyecto: Computación Serverless para el tratamiento de datos provenientes de dispositivos de IoT, cuya propuesta ha sido aprobada y está en desarrollo para el período 2020-2021, y se ha extendido un año más.

Asimismo el grupo de investigadores viene trabajando en proyectos relacionados con la computación distribuida y de alta performance desde hace más de 21 años. Como

continuación del proyecto anterior: Orquestación de Servicios para la Continuidad Edge al Cloud, se continúa el trabajo con investigadores de otras universidades, lo cual favorece notablemente a todas las instituciones participantes.

Introducción

La popularización de IoT y la masificación de las infraestructuras cloud durante el último tiempo se ha abierto un mundo de posibilidades para las aplicaciones HPC. Esto se debe a que los dispositivos IoT generan una gran cantidad de datos, los cuales se hace impráctico tratarlos con paradigmas tradicionales. Para lograr el procesamiento adecuado de estos datos con características de velocidad y tamaño importantes, se hace necesario prescindir de los paradigmas de programación tradicionales [1]. Es por ello que es necesario aplicar algoritmos que permitan aprovechar la escalabilidad de recursos de cómputo y procesamiento de datos [2] [3]. En este sentido se plantea como solución al procesamiento de datos provenientes del IoT, técnicas de computación de alta prestaciones (HPC) con el fin de aumentar la performance de procesamiento.

Los entornos HPC son ideales para resolver aplicaciones científicas, computacionalmente costosas con manejo de grandes cantidades de datos, a fin de lograr resultados en menor tiempo. Dadas estas características, estas arquitecturas son las mejores candidatas para procesar datos provenientes del IoT. Si bien, las arquitecturas HPC han evolucionado en pos de obtener mejores tiempos de respuesta para las aplicaciones, presentan el inconveniente del escalado, tanto vertical como horizontal, de recursos de cómputo. Es por ello que una alternativa es migrar al cloud [4].

Cloud Computing se ha caracterizado por ser una tecnología centrada en ofrecer cómputo bajo demanda como cualquier otro servicio. Esto es una ventaja para montar aplicaciones donde es necesario el procesamiento intensivo, tales como aquellas aplicaciones que procesen

y extraigan información de datos provenientes de dispositivos de IoT [5].

Los proveedores cloud alegan muchas ventajas en la migración de aplicaciones de HPC, como el acceso rápido a los recursos, costos más bajos y flexibilidad en la contratación y el aprovisionamiento de recursos. Un punto adicional es la seguridad, la cual, afirman es de alto nivel y en muchos casos muy difícil de implementar en la mayoría de los laboratorios, ya que tener personal de TI especializado en seguridad no es común [6].

Un aspecto más que lleva a la adopción del cloud como plataforma de despliegue de aplicaciones HPC es, mejorar la colaboración científica, es decir, facilitar la investigación colaborativa y la innovación; este aspecto ha sido el foco de este proyecto desde hace varios años al incorporar investigadores de otras universidades del país.

Otro tema es el potencial ahorro de tiempo que ofrece el cloud a los usuarios finales. Estos usuarios no necesitan preocuparse por actualizaciones de software, compatibilidad o parches de seguridad, pues todo esto es proporcionado de forma transparente.

Sin embargo, el cloud tiene dos grandes desventajas, la primera es la degradación de la performance de las aplicaciones al montarlas sobre arquitectura virtualizada, debido a que genera overhead en la contextualización de las máquinas virtuales; la segunda desventaja cuando se despliegan aplicaciones en el cloud, es que es responsabilidad de la organización mantener funcionando de forma correcta la infraestructura que se necesite para el despliegue de las aplicaciones, lo cual lleva a cargar costos sobre el presupuesto para su mantenimiento y soporte [7].

En este sentido, la aparición del Serverless Computing [8] logra que los desarrolladores no tengan que preocupar por el aprovisionamiento y escalado de la infraestructura, por lo que se pueden centrar en la lógica de sus aplicaciones. De esta forma es posible lograr la abstracción de la gestión de servidores (aprovisionamiento, configuración, escalado, etc.) para que los

usuarios, en este caso desarrolladores, puedan enfocarse en la lógica de sus aplicaciones.

HPC Serverless

Se ha mencionado en párrafos anteriores las ventajas y desventaja de migrar aplicaciones HPC al cloud. Sin embargo, el tiempo y el esfuerzo necesarios para configurar los recursos virtuales pueden ser mayores que el tiempo y el esfuerzo reales dedicados a hacer los cálculos. En contraposición, si se usa el paradigma serverless, será posible tener control más granular sobre el servicio prestado al dejar en manos del proveedor cloud la administración de la infraestructura.

En [9], se ha realizado un mapeo sistemático de 89 casos de uso donde se aplicó el paradigma serverless, para resolver problemas en su mayoría que se encuadran en HPC. Pero hay escasa información sobre una comparativa de performance entre las aplicaciones ejecutándose sobre paradigma serverless frente a las mismas aplicaciones ejecutándose sobre una infraestructura cloud tradicional, en la cual se pueda hacer un análisis de comportamiento para posteriormente decidir cuál es la mejor solución para ejecutar aplicaciones HPC

Otro punto ventajoso de serverless, pero que por ahora no es motivo de análisis de la presente propuesta, es el tiempo de desarrollo para las soluciones, dado que serverless permite que cada función pueda estar implementada en un lenguaje de programación diferente, en este caso se puede involucrar más desarrolladores a los proyectos.

Líneas de Investigación, Desarrollo e Innovación

La presente línea de investigación usa como punto de partida [10] y profundiza las tareas de investigación en base a [11], [12], [13], entre otras de los últimos años, en las cuales se han explorado y evaluado el rendimiento del uso de serverless en aplicaciones HPC. Si bien estos estudios demuestran que serverless es fácil de usar y económico, no se ha cuantificado su

efectividad sobre el enfoque convencional de aplicaciones corriendo en cloud.

En función de esta problemática el grupo de trabajo abordará esta problemática mediante la evaluación del comportamiento de las aplicaciones HPC ejecutándose en sobre infraestructura cloud tradicional contra serverless.

Para realizar esta tarea se realizará un análisis de los productos serverless que ofrece el mercado y de la factibilidad de ejecutar en ellos aplicaciones HPC que sean capaces de procesar datos provenientes de dispositivos IoT. Realizada esta tarea y seleccionado el proveedor cloud se realizará la evaluación en función de diferentes cargas de trabajo, diferentes tipos de datos, diferentes tipos de problemas y diferentes soluciones HPC tradicionales.

La metodología a seguir será experimental deductiva, lo cual permitirá analizar cómo se comportan las aplicaciones en diferentes entornos de ejecución.

Resultados y Objetivos

Resultados Obtenidos

Durante los últimos catorce años se trabajó en el área de Computación de Altas Prestaciones y distribuidas, en particular sobre análisis de diversas arquitecturas paralelas y distribuidas, tales como: Cloud Computing, Cluster de commodity, arquitecturas distribuidas y paralelas de bajo costo y fog computing.

El proyecto marco de esta línea de investigación, se inició hace dos años, el mismo tiene como temática principal a Serverless Computing. A partir de distintos análisis y debates en el grupo, se inició la línea de investigación del presente trabajo.

El grupo ha realizado varias publicaciones en esta área: ocho trabajos de investigación en Congresos y Jornadas, se realizaron tres publicaciones en revistas científicas y se transfirieron los resultados mediante conferencias en eventos científicos.

Se han aprobado dos tesinas de grado, se incorporó un becario de investigación categoría alumno y otra beca está en evaluación.

Objetivos

El objetivo del grupo de investigación es realizar un análisis de costos y performance en la ejecución de aplicaciones de HPC sobre serverless. Estas aplicaciones se plantean como aquellas capaces de manejar una gran cantidad de datos provenientes de dispositivos IoT, por lo que la variabilidad de la carga de trabajo también es un aspecto a evaluar.

De esta manera será posible evaluar la factibilidad de usar serverless para correr aplicaciones de HPC que posean cargas de trabajo variable.

Formación de Recursos Humanos

El equipo de trabajo de esta línea de investigación está compuesto de siete investigadores que figuran en este trabajo de las universidades Nacional de San Juan y Nacional de San Luis y dos alumnos de grado. Además, el proyecto marco donde se está desarrollando esta propuesta incluye a dos investigadores más de la Universidad Nacional de Salta, de la Universidad Champagnat y tres alumnos de grado.

Se está desarrollando una tesis doctoral sobre paralelismo híbrido y Big Data, una tesis de maestría en áreas afines y seis tesinas de grado en el área de Serverless computing, Concurrencia y Computación distribuida, en particular una sobre evaluación de la migración de HPC en el cloud a serverless y otra sobre bases de datos NewSQL.

Además se espera aumentar el número de publicaciones y se prevé la divulgación de varios temas investigados por medio de cursos de postgrado y actualización o publicaciones de divulgación y asesoramiento a empresas y otros organismos del estado.

Referencias

- [1] Farhan, L., Kharel, R., Kaiwartya, O., Quiroz-Castellanos, M., Alissa, A., & Abdulsalam, M. (2018, July). A concise review on Internet of Things (IoT)-problems, challenges and opportunities. In 2018 11Th International Symposium On Communication Systems, Networks & Digital Signal Processing (CSNDSP) (pp. 1-6). IEEE.
- [2] Medel, D., Murazzo, M. A., Molina, A. L., Sánchez, F., Cornejo, M., Rodríguez, N. R., ... & Piccoli, M. F. (2019). La Computación de Alta Performance como soporte a los sistemas altamente distribuidos. In XXI Workshop de Investigadores en Ciencias de la Computación (WICC 2019, Universidad Nacional de San Juan).
- [3] Barrionuevo, M., Escalante, J., Lopresti, M., Lucero, M., Miranda, N. C., Murazzo, M. A., & Piccoli, M. F. (2020). Solución de grandes problemas aplicando HPC multitecnología. In XXII Workshop de Investigadores en Ciencias de la Computación (WICC 2020, El Calafate, Santa Cruz).
- [4] de Souza Cimino, L., de Resende, J. E. E., Silva, L. H. M., Rocha, S. Q. S., de Oliveira Correia, M., Monteiro, G. S., ... & de Castro Lima, J. (2017, November). IoT and HPC integration: revision and perspectives. In 2017 VII Brazilian Symposium on Computing Systems Engineering (SBESC) (pp. 132-139). IEEE.
- [5] Biswas, A. R., & Giaffreda, R. (2014, March). IoT and cloud convergence: Opportunities and challenges. In 2014 IEEE World Forum on Internet of Things (WF-IoT) (pp. 375-376). IEEE.
- [6] Añel, J. A., Añel, J. A., Montes, D. P., Iglesias, J. R., & Romano. (2020). Cloud and Serverless Computing for Scientists. Springer International Publishing.
- [7] Malla, S., & Christensen, K. (2020). HPC in the cloud: Performance comparison of function as a service (FaaS) vs infrastructure as a service (IaaS). Internet Technology Letters, 3(1), e137.
- [8] Baldini, I., Castro, P., Chang, K., Cheng, P., Fink, S., Ishakian, V., ... & Suter, P. (2017). Serverless computing: Current trends and open

problems. In *Research advances in cloud computing* (pp. 1-20). Springer, Singapore.

[9] Eismann, S., Scheuner, J., Van Eyk, E., Schwinger, M., Grohmann, J., Herbst, N., ... & Iosup, A. (2020). A review of serverless use cases and their characteristics. *arXiv preprint arXiv:2008.11110*.

[10] Rodríguez, N. R., Murazzo, M. A., Medel, D., Parra, L., Molina, A. L., Sánchez, F., ... & Vargas, L. (2021). Procesamiento paralelo sobre arquitecturas serverless para tratamiento de datos provenientes del IoT. In *XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021, Chilecito, La Rioja)*.

[11] Niu, X., Kumanov, D., Hung, L. H., Lloyd, W., & Yeung, K. Y. (2019, September). Leveraging serverless computing to improve performance for sequence comparison. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics* (pp. 683-687).

[12] Spillner, J., Mateos, C., & Monge, D. A. (2017, September). Faaster, better, cheaper: The prospect of serverless scientific computing and hpc. In *Latin American High Performance Computing Conference* (pp. 154-168). Springer, Cham.

[13] Chard, R., Skluzacek, T. J., Li, Z., Babuji, Y., Woodard, A., Blaiszik, B., ... & Chard, K. (2019). Serverless supercomputing: High performance function as a service for science. *arXiv preprint arXiv:1908.04907*.

ALGORITMOS PARALELOS Y EVALUACION DE RENDIMIENTO EN PLATAFORMAS DE CÓMPUTO DE ALTAS PRESTACIONES

Marcelo Naiouf⁽¹⁾ , Armando De Giusti⁽¹⁾⁽²⁾ , Laura De Giusti⁽¹⁾⁽³⁾ , Franco Chichizola⁽¹⁾ , Victoria Sanz⁽¹⁾⁽³⁾ ,
Adrián Pousa⁽¹⁾ , Enzo Rucci⁽¹⁾⁽³⁾ , María José Basgall⁽¹⁾⁽²⁾ , Mariano Sánchez⁽¹⁾ , Manuel Costanzo⁽¹⁾ ,
Silvana Gallo⁽¹⁾⁽²⁾ , Emmanuel Frati⁽¹⁾⁽⁴⁾ , Adriana Gaudiani⁽⁵⁾ 

¹Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata
50 y 115, La Plata, Buenos Aires
Comisión de Investigaciones Científicas de la Pcia. de Buenos Aires (CIC)
526 e/ 10 y 11 La Plata Buenos Aires

²CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas

³CIC – Comisión de Investigación Científica de la Provincia de Buenos Aires

⁴Universidad Nacional de Chilecito

⁵Universidad Nacional de General Sarmiento

{mnaiouf, degiusti, ldgiusti, francoch, vsanz, apousa, erucci, sgallo, fefrati,
msanchez, mjbasgall}@lidi.info.unlp.edu.ar, agaudi@ungs.edu.ar

RESUMEN

El eje central de la línea de I/D es investigar en temas de cómputo paralelo y distribuido de alto desempeño, tanto en lo referido a los fundamentos como a la construcción, evaluación y optimización de las aplicaciones en arquitecturas multiprocesador. Se aplican los conceptos en problemas numéricos y no numéricos de cómputo intensivo y/o sobre grandes volúmenes de datos con el fin de obtener soluciones de alto rendimiento.

También incluye la construcción de ambientes para la enseñanza de la programación concurrente y paralela.

En la dirección de tesis de postgrado existe colaboración con el grupo HPC4EAS (High Performance Computing for Efficient Applications and Simulation) del Dpto. de Arquitectura de Computadores y Sistemas Operativos de la Universidad Autónoma de Barcelona; con el Departamento de Arquitectura de Computadores y Automática de la Universidad Complutense de Madrid; y con el grupo Soft Computing and Intelligent Information Systems (SCI2S) de la Universidad de Granada, entre otros.

Palabras clave: Cómputo paralelo y distribuido de altas prestaciones. Algoritmos paralelos y distribuidos. Arquitecturas multiprocesador. Ambientes de enseñanza.

CONTEXTO

La línea de I/D que se presenta es parte del Proyecto “Computación de Alto Desempeño: Arquitecturas, Algoritmos, Métricas de rendimiento y Aplicaciones en HPC, Big Data, Robótica, Señales y Tiempo Real.” del III-LIDI acreditado por el Ministerio de Educación, y de proyectos acreditados y subsidiados por la Facultad de Informática de la UNLP. Además, existe cooperación con Universidades de Argentina, Latinoamérica y Europa a través de proyectos acreditados por AECID, CyTeD, OEI y CIC y becas de Telefónica de Argentina. Asimismo, el III-LIDI forma parte del Sistema Nacional de Cómputo de Alto Desempeño (SNCAD).

1. INTRODUCCIÓN

El área de cómputo de altas prestaciones (HPC, High-Performance Computing) es clave dentro de las Ciencias de la Computación, debido al creciente interés por el desarrollo de soluciones a problemas con alta demanda computacional y de almacenamiento, produciendo transformaciones profundas en las líneas de I/D [1].

El rendimiento en este caso está relacionado con dos aspectos: las arquitecturas de soporte y los algoritmos que hacen uso de las mismas, y el desafío se centra en cómo aprovechar las prestaciones obtenidas a partir de la evolución de las arquitecturas físicas. En esta línea la mayor importancia está en los algoritmos paralelos y en los métodos utilizados para su construcción y análisis a fin de optimizarlos.

Uno de los cambios de mayor impacto ha sido el uso de manera masiva de procesadores con más de un núcleo (*multicore*), produciendo plataformas distribuidas híbridas (memoria compartida y distribuida) y generando la necesidad de desarrollar sistemas operativos, lenguajes y algoritmos que las usen adecuadamente. También creció la incorporación de placas aceleradoras a los sistemas multicore constituyendo plataformas paralelas de memoria compartida con paradigma de programación propio asociado como pueden ser las unidades de procesamiento gráfico (GPU, Graphic Processing Unit) de NVIDIA y AMD, los coprocesadores Xeon Phi de Intel [2] o los aceleradores basados en circuitos integrados reconfigurables (FPGAs, Field Programmable Gate Array) [3]. En la actualidad se comercializan placas de bajo costo como Raspberry PI [4] u Odroid [5] que poseen múltiples núcleos de baja complejidad y en algunos casos son procesadores multicore asimétricos (AMPs) con el mismo repertorio de instrucciones. Es de interés estudiar como explotar el paralelismo en estos dispositivos para mejorar el rendimiento y/o consumo energético de las aplicaciones [6], así como las características de scheduling en los mismos [7]. Asimismo, los entornos de computación cloud introducen

un nuevo foco desde el punto de vista del HPC, brindando un soporte “a medida” sin la necesidad de adquirir el hardware.

La creación de algoritmos paralelos en arquitecturas multiprocesador no es un proceso directo [8]. El costo puede ser alto en términos del esfuerzo de programación y el manejo de la concurrencia adquiere un rol central en el desarrollo. Si bien en las primeras etapas el diseñador de una aplicación paralela puede abstraerse de la máquina sobre la que ejecutará el algoritmo, para obtener buen rendimiento debe tenerse en cuenta la plataforma de destino. En las máquinas multiprocesador, se deben identificar las capacidades de procesamiento, interconexión, sincronización y escalabilidad. La caracterización y estudio de rendimiento del sistema de comunicaciones es de interés para la predicción y optimización de performance, así como la homogeneidad o heterogeneidad de los procesadores [9].

Muchos problemas algorítmicos se vieron impactados por los multicore y clusters de multicore. A partir de incorporar varios chips multicore dentro de un nodo y conectar múltiples nodos vía red, se puede crear una arquitectura NUMA, de modo que los cores en un chip compartan memoria principal, y puedan acceder remotamente a la memoria dedicada de otro chip, aunque ese acceso sea más costoso, surgiendo así varios niveles de comunicación. Esto impacta sobre el desarrollo de algoritmos que aprovechen adecuadamente las arquitecturas, y motiva el estudio de performance en sistemas híbridos. Además, es necesario estudiar la utilización de lenguajes y bibliotecas ya que aún no se cuenta con un standard, aunque puede mencionarse el uso de los tradicionales MPI, OpenMP y Pthreads [10][11] o los más recientemente explorados UPC, Chapel y Titanium del modelo PGAS [12].

La combinación de arquitecturas de múltiples núcleos con aceleradores dio lugar a plataformas híbridas con diferentes características. Más allá del acelerador utilizado, la programación de estas plataformas representa un desafío. Para lograr aplicaciones de alto rendimiento, los programadores enfrentan dificultades como: estudiar características específicas de cada arquitectura y aplicar técnicas de programación y optimización particulares de cada una, lograr un balance de carga adecuado entre los dispositivos de procesamiento y afrontar la ausencia de estándares para este tipo de sistemas.

Por otra parte, los avances en las tecnologías de virtualización han llevado a que Cloud Computing sea una alternativa a los tradicionales sistemas de cluster [13]. El uso de cloud para HPC presenta desafíos atractivos, brindando un entorno reconfigurable dinámicamente sin la necesidad de adquirir hardware, y es una excelente plataforma para testear escalabilidad de algoritmos aunque queda mucho por hacer en cuanto al diseño, lenguajes y programación

Métricas de evaluación del rendimiento y balance de carga

La diversidad de opciones vuelve complejo el análisis de performance de los Sistemas Paralelos, ya que los ejes sobre los cuales pueden compararse dos sistemas son varios. Existe un gran número de métricas para evaluar el rendimiento, siendo las tradicionales: tiempo de ejecución, speedup, eficiencia. Por su parte, la *escalabilidad* permite capturar características de un algoritmo paralelo y la arquitectura en que se lo implementa. Posibilita testear la performance de un programa sobre pocos procesadores y predecirla en un número mayor, así como caracterizar la cantidad de paralelismo inherente en un algoritmo.

Un aspecto de interés que se ha sumado como métrica, a partir de las plataformas con gran cantidad de procesadores, es el del *consumo* y la *eficiencia energética* [14]. Muchos esfuerzos están orientados a tratar el consumo como eje de I/D, como métrica de evaluación, y también a la necesidad de metodologías para medirlo.

El objetivo principal del cómputo paralelo es reducir el tiempo de ejecución haciendo uso eficiente de los recursos. El *balance de carga* es un aspecto central y consiste en, dado un conjunto de tareas que comprenden un algoritmo y un conjunto de procesadores, encontrar el mapeo (asignación) de tareas a procesadores tal que cada una tenga una cantidad de trabajo que demande aproximadamente el mismo tiempo, y esto es más complejo si hay heterogeneidad. Dado que el problema general de mapping es *NP-completo*, pueden usarse enfoques que dan soluciones subóptimas aceptables. Las técnicas de planificación a nivel micro (dentro de cada procesador) y macro (en un cluster) deben ser capaces de obtener buen balance de carga. Existen técnicas estáticas y dinámicas cuyo uso depende del conocimiento que se tenga sobre las tareas de la aplicación.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

- Investigar en temas de cómputo paralelo y distribuido de alto desempeño, en lo referido a los fundamentos y a la construcción y evaluación de las aplicaciones. Esto incluye los problemas de software asociados con el uso de arquitecturas multiprocesador:
 - Lenguajes, modelos y paradigmas de programación paralela (puros e híbridos a distintos niveles).
 - Asignación de procesos a procesadores optimizando el balance de la carga de procesamiento.
 - Métricas de evaluación de complejidad y rendimiento: speedup, eficiencia, escalabilidad, consumo energético, costo de programación.
- Construir, evaluar y optimizar soluciones utilizando algoritmos concurrentes, paralelos y distribuidos sobre diferentes plataformas de software y arquitecturas con múltiples procesadores:
 - Arquitecturas de trabajo homogéneas, heterogéneas e híbridas: multicores, clusters,

GPU, Xeon Phi, FPGA, placas de bajo costo y entornos cloud.

- Aplicar los conceptos en problemas numéricos y no numéricos de cómputo intensivo y/o sobre grandes volúmenes de datos (aplicaciones científicas, búsquedas, simulaciones, imágenes, realidad virtual y aumentada, bioinformática, big data, n-body).
- Analizar y desarrollar ambientes para la enseñanza de programación concurrente y paralela.
 - Caracterizar diferentes modelos de arquitecturas paralelas.
 - Representar distintos modelos de comunicación/sincronización.
 - Definir métricas de evaluación de rendimiento y eficiencia energética.

3. RESULTADOS OBTENIDOS/ESPERADOS

- Desarrollar y optimizar algoritmos paralelos sobre diferentes modelos de arquitectura. En particular, en aplicaciones numéricas y no numéricas de cómputo intensivo y tratamiento de grandes volúmenes de datos.
- Estudiar y comparar los lenguajes sobre las plataformas multiprocesador para diferentes modelos de interacción entre procesos.
- Investigar la paralelización en plataformas que combinan clusters, multicore y aceleradores. Comparar estrategias de distribución de trabajo teniendo en cuenta las diferencias en potencias de cómputo y comunicación, dependencia de datos y memoria requerida.
- Evaluar la performance (speedup, eficiencia, escalabilidad, consumo energético) de las soluciones propuestas. Analizar el rendimiento de soluciones paralelas a problemas con diferentes características (dependencia de datos, relación cómputo / comunicación, memoria requerida).
- Mejorar y adecuar las técnicas disponibles para el balance de carga (estático y dinámico) entre procesos a las arquitecturas consideradas.

En este marco, pueden mencionarse los siguientes resultados:

- Para la experimentación se han utilizado y analizado diferentes arquitecturas homogéneas o heterogéneas, incluyendo multicores, cluster de multicores (con 128 núcleos), GPU y cluster de GPU, Xeon Phi y FPGA.
- Se experimentó la paralelización en arquitecturas híbridas, con el objetivo de estudiar el impacto del mapeo de datos y procesos, así como de los lenguajes y librerías.
- Respecto de las aplicaciones y temas estudiados, se trabajó fundamentalmente con los siguientes problemas:

➤ **Aceleración de aplicaciones con cómputo colaborativo CPU-GPU.** Las computadoras comerciales actuales incluyen decenas de cores y al menos una GPU. El uso de ambas unidades de procesamiento de forma colaborativa puede mejorar significativamente el rendimiento de una aplicación. Sin

embargo, esto supone un desafío para los programadores ya que dichas unidades difieren en arquitectura, modelo de programación y rendimiento. En [15] se propuso un modelo híbrido para estructurar código a ser ejecutado sobre un sistema heterogéneo con múltiples cores y 1 GPU (utilizando todos los recursos disponibles). Utilizando este modelo se desarrolló un algoritmo paralelo de pattern matching para sistemas heterogéneos CPU-GPU [16]. Los resultados revelaron que este algoritmo supera en rendimiento a trabajos previos, desarrollados para sistemas multicore y GPUs, para datos de tamaño considerable. En [17] se presentó una solución al problema de pattern matching que aprovecha toda la potencia computacional de los procesadores Intel Xeon Phi KNL 7230 mediante el uso de SIMD y paralelismo de hilos. Se mostró que el algoritmo propuesto alcanza aceleraciones significativas. En trabajos futuros interesa investigar sobre el cómputo colaborativo incluyendo este tipo de arquitecturas.

➤ **Alineamiento de secuencias biológicas.** Esta operación consiste en comparar dos o más secuencias biológicas, como pueden ser las de ADN o las de proteínas, y resulta fundamental en investigaciones de la bioinformática y la biología molecular. El algoritmo de Smith-Waterman es considerado el método de alineamiento más preciso. Desafortunadamente, este algoritmo resulta costoso debido a su complejidad computacional cuadrática mientras que la situación se agrava aún más a causa del crecimiento exponencial de datos biológicos en los últimos años [18]. El reciente surgimiento de aceleradores en HPC (GPU, Xeon Phi, FPGA, entre otros) da la oportunidad de acelerar los alineamientos sobre hardware comúnmente disponible a un costo accesible, como se ha mostrado en [19][20][21][22]. A futuro, interesa explorar las fortalezas y debilidades del uso de nuevas tecnologías de software para arquitecturas paralelas como, por ejemplo, oneAPI [23].

➤ **Cálculo de los caminos mínimos.** Es uno de los problemas básicos y de mayor antigüedad de la teoría de grafos teniendo aplicación en el dominio de las comunicaciones, del ruteo de tráfico, de la bioinformática, entre otros. El algoritmo de Floyd-Warshall (FW) permite computar la distancia mínima entre todos los pares de un grafo. Además de poseer una alta demanda de ancho de banda, FW resulta costoso computacionalmente al ser $O(n^3)$. Se desarrollaron implementaciones optimizadas para dos arquitecturas HPC recientes -como son Intel Xeon Phi KNL y NVIDIA Pascal- y se analizó comparativamente su rendimiento y eficiencia energética (teórica) en diferentes escenarios. Como trabajo futuro, interesa incorporar otros modelos de las arquitecturas elegidas al estudio para robustecer el análisis [24].

➤ **Simulaciones utilizando modelos basados en agentes.** El objetivo de esta línea son las simulaciones basadas en agentes sobre infraestructuras de altas prestaciones con modelos más cercanos a la realidad que ayuden a la toma de decisiones a científicos de otras disciplinas, y no expertos en el área de la informática, (biología, ecología, física, etc.). Para ello, y por ser modelos complejos que necesitan mucha potencia de cómputo, resulta imprescindible el uso de soluciones de HPC con el fin de lograr tiempos de respuesta reducidos para entornos complejos. Se busca desarrollar algoritmos y simuladores como soluciones HPC que sean eficientes y escalables y por ello una cuestión importante, tratada en esta línea de trabajo, es lograr que además de estas premisas, la simulación sea realizada con el menor consumo posible de energía. Dado que son simulaciones con un gran número de ejecuciones por escenario, no solo es necesario que la solución tenga buenas prestaciones sino que sea posible predecir, mediante un modelo energético, la energía necesaria para llevar a cabo diferentes escenarios de simulación.

➤ **Simulación de N cuerpos computacionales con atracción gravitacional.** Su propósito es aproximar en forma numérica la evolución de un sistema de cuerpos en el que cada uno interactúa con todos los restantes. El uso más conocido de esta simulación quizás sea en la astrofísica, donde cada cuerpo representa una galaxia o una estrella particular que se atraen entre sí debido a la fuerza gravitacional. Si bien existen diferentes métodos para procesar la simulación de los N cuerpos, en todos los casos se requiere alta demanda computacional. Se estudió la paralelización de la versión directa de esta simulación sobre diferentes lenguajes no convencionales en el ámbito de HPC, en particular Python y Rust. Se focalizó en cómo diferentes técnicas de optimización logran mejorar el rendimiento y en su comparación con un lenguaje tradicional como es C, considerando no sólo el rendimiento sino también el esfuerzo de programación. A futuro, interesa extender el estudio incorporando otras aplicaciones y arquitecturas multicore [25][26].

➤ **Problemas de optimización de simulación de sistemas dinámicos complejos mediante heurísticas.** La búsqueda de un conjunto de parámetros de entrada que optimicen el funcionamiento de un simulador de un sistema físico es un proceso de alto costo computacional que puede considerarse intratable y requiere de heurísticas que permitan disminuir el tiempo de ejecución. En los nuevos avances realizados en esta línea se aprovechó la continuidad en los valores de los parámetros físicos distribuidos sobre el dominio del sistema, de manera de realizar búsquedas de parámetros ajustados sobre espacios de búsquedas de tamaño mucho más reducidos que los utilizados en una primera etapa del trabajo. Esta metodología mucho más eficiente se puede extender a otros simuladores de fenómenos

físicos y en particular con simuladores de inundaciones de ríos con las que se llevaron adelante las experiencias. Las experiencias se pueden correr de manera colaborativa beneficiándose ampliamente del uso de plataformas de clusters de procesadores [27] [28].

➤ **Ambientes para la enseñanza de concurrencia.** Se desarrolló el entorno CMRE para la enseñanza de programación concurrente y paralela a partir de cursos iniciales en carreras de Informática. Incluye un entorno visual que representa una ciudad en la que pueden definirse varios robots que interactúan. Combina aspectos de memoria compartida y distribuida mediante instrucciones para bloquear y liberar esquinas de la ciudad y el concepto de pasaje de mensajes a través de primitivas de envío y recepción. Además, se incluyen los conceptos de heterogeneidad (diferentes velocidades de los robots) y consumo energético [29]. Se ha integrado con el uso de robots físicos (Lego Mindstorm 3.0) que ejecutan en tiempo real las mismas instrucciones que los robots virtuales y se comunican con el entorno mediante bluetooth [30]. Se ha ampliado para incorporar conceptos básicos de computación en la nube (Cloud Computing) [30]. Actualmente, se está desarrollando una nueva herramienta para la enseñanza de programación concurrente en cursos avanzados. Su objetivo principal es visualizar los conceptos de sincronización y comunicación entre procesos.

➤ **Aplicaciones en Big Data.** En los últimos años, los escenarios de grandes volúmenes de datos (Big Data) son cada vez más comunes debido a la constante generación de datos a partir de fuentes como sensores o Internet, por ejemplo. Para poder trabajar con conjuntos de datos Big Data, se requiere de cómputos de altas prestaciones y de herramientas software específicas capaces de procesar significativos tamaños de datos en tiempos razonables, y que puedan ser ejecutados de manera paralela y distribuida.

Una de las herramientas más populares para el procesamiento de Big Data es Apache Spark, la cual presenta varias características deseables que la hacen ser ampliamente elegida. El uso intensivo de memoria RAM, los mecanismos eficientes de tolerancia a fallos, las distintas estructuras de datos que provee altamente optimizadas, la amplia variedad de librerías que habilitan utilizar algoritmos de Machine Learning, procesamiento en Streaming, entre otros, son algunas de sus características más relevantes.

En esta línea de trabajo, se están desarrollando nuevas técnicas para el preprocesamiento de datos para problemas de clasificación en Big Data, utilizando el framework Apache Spark sobre Cómputo de Altas Prestaciones. Como parte del trabajo realizado se presenta una propuesta metodológica para la reducción dual (reducción de instancias y de características) de un conjunto de datos tabulares que representa un problema de clasificación [31].

Se trata de un diseño sencillo y escalable, capaz de analizar la reducción de datos de forma vertical y horizontal, recibiendo la visión del experto en datos mediante el establecimiento de dos simples condiciones: un umbral de calidad predictiva asociado al modelo obtenido a partir del conjunto reducido, y un rango de porcentajes de reducción a comprobar. Además, su implementación utiliza operaciones paralelas y utilidades totalmente optimizadas proporcionadas por Apache Spark.

Esta línea de trabajo se está llevando a cabo en colaboración con el Dr. Alberto Fernández Hilario del grupo de investigación "Soft Computing and Intelligent Information Systems" (SCI2S) de la Universidad de Granada, España.

➤ **Cifrado de grandes volúmenes de datos.** Hoy en día, la cantidad de datos sensibles que se generan para ser almacenados y/o transmitidos a través de la red aumenta constantemente. Para proteger los datos confidenciales de amenazas potenciales, se utilizan estrategias de encriptación. Además, el tiempo involucrado en el cifrado de datos está directamente relacionado con la cantidad de datos que se cifrarán y puede ser significativo. Para reducir el tiempo de cifrado es natural recurrir a soluciones de encriptación paralelas, que explotan todo el poder computacional proporcionado por las arquitecturas emergentes. AES (Advanced Encryption Standard) es uno de los algoritmos de cifrado más utilizados y el gobierno de los Estados Unidos lo considera lo suficientemente seguro como para proteger la información nacional. Hay varias implementaciones de AES, tanto en hardware como en software. En [32] se presenta una comparación de rendimiento de una solución AES basada en hardware para CPU multinúcleo con el de otras dos soluciones AES basadas en software para CPU multinúcleo y GPU, respectivamente. El primero se implementa con las nuevas instrucciones de Intel AES y el segundo con la biblioteca OpenSSL. Los resultados revelan que utilizar cómputo paralelo para el proceso de cifrado reduce significativamente el tiempo de ejecución respecto a una solución secuencial. Los resultados también muestran que el mayor rendimiento se alcanza utilizando una solución multicore con AES basado en hardware. Sin embargo, la solución por software utilizando 2 GPUs puede ser una alternativa competitiva a la solución multicore por hardware cuando se tienen pocos núcleos o una CPU que no soporta AES.

4. FORMACIÓN DE RECURSOS HUMANOS

Dentro de la temática de la línea de I/D se concluyó 1 tesis doctoral, 1 Trabajo Final de Especialización y 2 Tesinas de Grado de Licenciatura. Se encuentran en curso en el marco del proyecto 3 tesis doctorales, 2 de maestría, 2 trabajos de Especialización y 4 Tesinas de grado.

Se participa en el dictado de las carreras de Doctorado en Cs. Informáticas y Magíster y Especialización en Cómputo de Altas Prestaciones de la Facultad de Informática UNLP, por lo que potencialmente pueden generarse más Tesis y Trabajos Finales.

Hay cooperación con grupos de otras Universidades del país y del exterior, y tesistas de diferentes Universidades realizan su trabajo con el equipo del proyecto.

5. BIBLIOGRAFÍA

- [1]. Giles MB, Reguly I. "Trends in high-performance computing for engineering calculations". *Phil.Trans.R.Soc.A* 372: 20130319. 2014. <http://dx.doi.org/10.1098/rsta.2013.0319>
- [2]. Jeffers, James; Reinders, James. "Intel Xeon Phi Coprocessor High Performance Programming". Morgan Kaufmann. 2013.
- [3]. Sean Settle. "High-performance Dynamic Programming on FPGAs with OpenCL". *IEEE High Performance Extreme Computing Conf (HPEC '13)*. 2013. <https://doi.org/10.1016/j.physa.2017.04.159>.
- [4]. Raspberry PI. <https://www.raspberrypi.org/> Accedido 21 de Marzo de 2016.
- [5]. Odroid <http://www.hardkernel.com> Accedido 21 de Marzo de 2016.
- [6]. Annamalai A., Rodrigues R., Koren I., Kundu S., "Dynamic Thread Scheduling in Asymmetric Multicores to Maximize Performance-per-Watt," 2012 IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum, pp. 964-971, 2012 IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum, 2012.
- [7]. Juan Carlos Saez, Adrian Pousa, Daniel Chaver, Fernando Castro, Manuel Prieto Matias. "ACFS: A Completely Fair Scheduler for Asymmetric Single-ISA Multicore Systems". *ACM SAC 2015 (The 30TH ACM/SIGAPP Symposium on applied computing)*. 2015.
- [8]. McCool, Michael. "Structured Parallel Programming: Patterns for Efficient Computation", Morgan Kaufmann, 2012
- [9]. De Giusti L, Naiouf M., Chichizola F., Luque E., De Giusti A. "Dynamic Scheduling in Heterogeneous Multiprocessor Architectures. Efficiency Analysis". *Computer Science and Technology Series – XV Argentine Congress of Computer Science Selected Papers*. La Plata (Buenos Aires): Editorial de la Universidad de La Plata (edulp). 2010. p85 - 95. isbn 978-950-34-0684-7.
- [10]. Chapman, B., Jost, G. & Van der Pas. "Using OpenMP – Portable Shared Memory Parallel Programming". (2008). UK: MIT Press.
- [11]. Hager, G. & Wellein, G. "Introduction to HPC for Scientists and Engineers". (2011) EEUU: CRC Press.
- [12]. De Wael, M; Marr, S; De Fraine, B; Van Cutsem, T; De Meuter, W. "Partitioned Global Address Space Languages". *ACM Computing Surveys (CSUR)* 47 (4), 2015.

- [13]. Amazon Elastic Compute Cloud (Amazon EC2). <http://aws.amazon.com/es/ec2/>. Febrero 2013.
- [14]. Balladini J., Rucci E., De Giusti A., Naiouf M., Suppi R., Rexachs D., Luque E. "Power Characterisation of Shared-Memory HPC Systems". Computer Science & Technology Series – XVIII Argentine Congress of Computer Science Selected Papers. ISBN 978-987-1985-20-3. Pp. 53-65. EDULP, La Plata (Argentina), 2013
- [15]. V. Sanz, A. Pousa, M. Naiouf, and A. De Giusti, "Accelerating Pattern Matching with CPU-GPU Collaborative Computing", Proceedings of the International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP 2018), ISBN: 978-3-030-05051-1, págs. 310-322, doi. https://doi.org/10.1007/978-3-030-05051-1_22, 2018.
- [16]. Sanz V., Pousa A., Naiouf M., De Giusti A. (2020) Efficient Pattern Matching on CPU-GPU Heterogeneous Systems. In: Wen S., Zomaya A., Yang L. (eds) Algorithms and Architectures for Parallel Processing. ICA3PP 2019. Lecture Notes in Computer Science, vol 11944. Springer, Cham.
- [17]. Sanz V., Pousa A., Naiouf M., De Giusti A. (2020) Accelerating Pattern Matching on Intel Xeon Phi Processors. In: Qiu M. (eds) Algorithms and Architectures for Parallel Processing. ICA3PP 2020. Lecture Notes in Computer Science, vol 12452. Springer, Cham. https://doi.org/10.1007/978-3-030-60245-1_18.
- [18]. Enzo Rucci, Armando De Giusti, Marcelo Naiouf, Carlos García Sanchez, Guillermo Botella Juan, Manuel Prieto-Matías. "State-of-the-art in Smith-Waterman Protein Database Search". Big Data Analytics in Genomics. Ka-Chun Wong (Editor). ISBN: 978-3-319-41278-8 (print) 978-3-319-41279-5 (online), Springer, págs. 197-223, 2016.
- [19]. E. Rucci, C. Garcia, G. Botella, A. De Giusti, M. Naiouf, and M. Prieto-Matías. "Accelerating Smith-Waterman Alignment of Long DNA Sequences with OpenCL on FPGA". En: Bioinformatics and Biomedical Engineering. IWBBIO 2017. Lecture Notes in Computer Science, vol 10209., ISBN: 978-3-319-56154-7, Springer, Cham, págs. 500-511, 2017.
- [20]. E. Rucci, C. Garcia, G. Botella, A. De Giusti, M. Naiouf, and M. Prieto-Matías. "SWIFOLD: Smith-Waterman Implementation on FPGA with OpenCL for Long DNA Sequences". BMC Systems Biology. ISSN: 1752-0509. In press. 2018.
- [21]. E. Rucci, C. Garcia, G. Botella, A. De Giusti, M. Naiouf, and M. Prieto-Matías. "First Experiences Accelerating Smith-Waterman on Intel's Knights Landing Processor". Algorithms and Architectures for Parallel Processing. ICA3pp 2017. Lecture Notes in Computer Science, vol 10393, ISBN: 978-3-319-65482-9, Springer International Publishing, págs. 569-579, 2017.
- [22]. Enzo Rucci, Carlos García, Guillermo Botella, Armando De Giusti, Marcelo Naiouf and Manuel Prieto-Matías. "SWIMM 2.0: enhanced Smith-Waterman on Intel's Multicore and Manycore architectures based on AVX-512 vector extensions". International Journal of Parallel Programming, Springer US, 2018. DOI: <https://doi.org/10.1007/s10766-018-0585-7>
- [23]. one API Programming Model, Disponible en www.oneapi.org, Accedido el 15/02/2021.
- [24]. M. Costanzo, E. Rucci, U. Costi, F. Chichizola, and M. Naiouf, "Comparison of HPC Architectures for Computing All-Pairs Shortest Paths. Intel Xeon Phi KNL vs NVIDIA Pascal". En: Computer Science – CACIC 2020. Revised Selected Papers., Springer International Publishing, págs. 37-49, doi. 10.1007/978-3-030-75836-3_3, 2021.
- [25]. M. Costanzo, E. Rucci, M. Naiouf, and A. D. Giusti, "Performance vs Programming Effort between Rust and C on Multicore Architectures: Case Study in N-Body", Proceedings of 2021 XLVII Latin American Computing Conference (CLEI), ISBN: 978-1-66549-503-5, págs. 1-10, doi. 10.1109/CLEI53233.2021.9640225, 2021.
- [26]. A. Milla and E. Rucci, "Acelerando código científico en Python usando Numba", Actas del XXVII Congreso Argentino de Ciencias de la Computación (CACIC 2021), ISBN: 978-987-633-574-4, págs. 72-82, 2021.
- [27]. Trigila M., Gaudiani A., Luque E. (2018) "Agile Tuning Method in Successive Steps for a River Flow Simulator". In: Shi Y. et al. (eds) Computational Science – ICCS 2018. ICCS 2018. Lecture Notes in Computer Science, vol 10862. Springer, Cham
- [28]. Trigila, Mariano, Gaudiani, Adriana, Luque, Emilio. "Adjustment of a simulator of a complex dynamic system with emphasis on the reduction of computational resources". Actas del Workshop. p. 142-147. 2018.
- [29]. J. Castro, L. D. Giusti, G. Gorga, M. Sánchez, and M. Naiouf, "ECMRE: Extended Concurrent Multi Robot Environment". Computer Science – CACIC 2017. Communications in Computer and Information Science, vol 790, ISBN: 978-3-319-75213-6 978-3-319-75214-3, Springer, Cham, págs. 285-294. 2018.
- [30]. L. C. De Giusti, F. Chichizola, S. Rodríguez Eguren, M. Sanchez, J. M. Paniego, A. E. De Giusti. "Introduciendo conceptos de Cloud Computing utilizando el entorno CMRE". Proceedings del XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016) – Workshop de Innovación en Educación en Informática. Octubre 2016. Pp 1357-1365.
- [31]. Basgall, M. J., Naiouf, M., & Fernández, A. (2021). FDR2-BD: A Fast Data Reduction Recommendation Tool for Tabular Big Data Classification Problems. Electronics, 10(15), 1757. <https://doi.org/10.3390/electronics10151757>.
- [32]. Sanz V., Pousa A., Naiouf M., De Giusti A. (2021) Comparison of Hardware and Software Implementations of AES on Shared-Memory Architectures. In: Naiouf M., Rucci E., Chichizola F., De Giusti L. (eds) Cloud Computing, Big Data & Emerging Topics. JCC-BD&ET 2021. Communications in Computer and Information Science, vol 1444. Springer, Cham. https://doi.org/10.1007/978-3-030-84825-5_5.

Tecnología HPC en la UNdeC como motor de ciencia

Fernando Emmanuel FRATI¹, José TEXIER¹, Paula Cecilia RIVERA¹, Jonathan ALVAREZ¹,
 Fernanda CARMONA¹, Patricia FIGUEROLA¹, Francisco FRATI¹, Sebastián GUIDET¹, Roberto
 MILLON¹, Raul MORALEJO¹, Matías PEREZ¹, Emmanuel PORTUGAL¹, Donna RATTALINO¹, Alberto
 Eduardo RIBA¹, Daniel ROBINS¹, Mara ROVERO¹, Javier RUITTI¹, Jorge TEJADA¹, Jusmeidy
 ZAMBRANO¹,
 Carlos Esteban GRAFFIGNA¹, Javier BALLADINI², Enzo RUCCI³, Ruber HERNÁNDEZ GARCÍA⁴

¹ Universidad Nacional de Chilecito

9 de julio 22, Chilecito, La Rioja, Argentina

{fefrati, jtexier, privera, jalvarez, fbcarmona, pfiguerola, flfrati,
 sguidet, rmillon, rmoralejo, mperez, eportugal, drattalino, ariba,
 drobins, mrovero, jruiitti, jtejada, jzambrano,
 cgraffigna}@undec.edu.ar,

² Universidad Nacional del Comahue

javier.balladini@fi.uncoma.edu.ar,

³ Instituto de Investigación en Informática III-LIDI, Universidad Nacional de La Plata

erucci@lidi.info.unlp.edu.ar,

⁴ Laboratorio de Investigaciones Tecnológicas en Reconocimiento de Patrones (LITRP),

Universidad Católica de Maule

rhernandez@ucm.cl

RESUMEN

En la actualidad, para abordar problemas de mayor tamaño y complejidad, los estudios de ciencia básica y aplicada utilizan Computación de Altas Prestaciones (HPC - High Performance Computing). El HPC permite mejorar la capacidad, velocidad y precisión en el procesamiento de datos. Con el proyecto que da origen a este trabajo se abordan seis estudios desde la perspectiva del HPC, para explorar los aspectos centrales del paralelismo aplicado desde las Ciencias de la Computación en otras disciplinas.

Algunos de estos estudios se realizan exclusivamente en la Universidad Nacional de Chilecito, mientras que otros son en cooperación con otras instituciones nacionales y extranjeras. Entre estos, tres formalizan trabajos finales de postgrado. En todos los casos, el HPC es abordado a través de un proceso metodológico organizado para:

- Consolidar una infraestructura de experimentación, desarrollo y producción de soluciones a problemas de HPC
- Desarrollar las capacidades científico-tecnológicas del equipo

- Fomentar la vinculación y transferencia con los sectores académico, social y productivo
- Cada problema abordado reúne entre sus integrantes investigadores especialistas en la disciplina del estudio, investigadores de Ciencias de la Computación y estudiantes en sus últimos años de formación de grado. Con esto, se está consolidando un grupo de investigación, desarrollo y transferencia que generará oportunidades de formación de recursos humanos, proveerá de servicios a la comunidad en el área de estudio y potenciará los vínculos de cooperación con otras instituciones.

Palabras clave: HPC, cómputo paralelo, aplicaciones, interdisciplinariedad.

CONTEXTO

La línea de investigación presentada es parte del proyecto “Software y aplicaciones en Computación de Altas Prestaciones” fue aprobado en la convocatoria a proyectos de Investigación y Desarrollo 2018 de la Secretaría de Ciencia y Tecnología de la UNdeC, y se encuentra en ejecución desde junio de 2019. Además, en 2018 la UNdeC destinó fondos de PROMINF para la adquisición de 12 PC con procesadores i7 y 8GB RAM, 5 de las cuales

están equipadas con placas de video NVIDIA GTX 1060 para el “Laboratorio de Sistemas Paralelos”. También, con el objeto de desarrollar las capacidades en HPC de la UNDeC, a fines de 2019 se adquirió un servidor Dell PowerEdge R740, equipado con 2 Xeon Platinum 8176 (56 núcleos físicos, 112 threads en total), 256 GB de RAM y 2 GPGPU NVIDIA Quadro P4000 (financiado a través del “Plan de mejoramiento de la función de I+D+i” - MINCyT), actualmente en funcionamiento. Finalmente, se destinaron fondos del PROMINF para financiar parcialmente la formación de postgrado de docentes afines a la disciplina, 20 de los cuales están realizando la *Especialización en inteligencia de datos orientada a big data* (acreditada por CONEAU) de la Facultad de Informática de la UNLP. Durante el desarrollo del proyecto se iniciaron y consolidaron colaboraciones con las universidades UNLP (Argentina), UCM (Chile) y UNCo (Argentina). Estas iniciativas permiten consolidar una infraestructura de experimentación, desarrollo y producción de soluciones a problemas de HPC, como así también recursos humanos formados para aprovecharla.

1. INTRODUCCIÓN

La informática tiene su origen en la necesidad de los distintos sectores de la sociedad de conseguir mayor velocidad, confiabilidad y precisión para resolver sus problemas. Sin embargo, la capacidad de solución a un problema dado encuentra su límite en los tiempos requeridos por sus algoritmos. Superar ese límite requiere que el problema sea abordado mediante cómputo paralelo. Normalmente, esto implica estudiar tres aspectos clave: hardware, aplicaciones y software.

Durante décadas, la industria respondió a la creciente demanda de mayor poder computacional incrementando exponencialmente el rendimiento de los procesadores [1]. Sin embargo, esta forma de obtener mayor poder de cómputo encontró barreras físicas, limitando el rendimiento de los microprocesadores y los sistemas en general [2]. Desde el año 2005, el escalado tecnológico se

ha venido aprovechando para aumentar el número de cores dentro del chip, dando lugar a una importante variedad de arquitecturas (multicores, commodity clusters, GPGPU y Cloud) [3], [4].

No obstante, reducir los tiempos de procesamiento y obtener la mayor eficiencia de ese hardware requiere el diseño y desarrollo de algoritmos paralelos [5]. Transformar un algoritmo secuencial en uno paralelo no es trivial. En general los procesos concurrentes necesitan algún mecanismo para comunicar resultados parciales entre sí. Dependiendo de la arquitectura de cómputo, se consigue a través del uso de variables compartidas o del paso de mensajes entre procesos. Una transformación ‘implícita’ o transparente es deseable, pero el costo es una pérdida importante de rendimiento [6]. En su lugar, el programador recurre a librerías estándares para expresar explícitamente el paralelismo: OpenMP, Pthreads, CUDA, OpenCL, MPI [7].

Las suposiciones de orden de ejecución entre instrucciones heredadas del modelo de programación secuencial ya no son válidas, obligando al programador a utilizar algún mecanismo de sincronización para garantizar estados consistentes del programa. En este contexto, la correctitud de los algoritmos es más difícil de garantizar que en la computación serial. Frecuentemente ocurren errores al sincronizar los procesos, dando lugar a deadlocks, condiciones de carrera, violaciones de orden, violaciones de atomicidad simple y violaciones de atomicidad multivariable, requiriendo el uso de herramientas de depuración específicas [8]. A diferencia de soluciones secuenciales donde existe un modelo teórico que permite estimar el desempeño de los programas antes de escribirlos, la evaluación del sistema paralelo (software y hardware) requiere la definición de distintas métricas: tiempo de ejecución, speedup, eficiencia y overhead. Las soluciones paralelas están tan estrechamente vinculadas con el hardware subyacente que dificultan enormemente conseguir portabilidad de rendimiento [9].

Existen muchos aspectos que requieren ser tomados en cuenta al diseñar la solución

paralela: tamaño del problema, división de datos o tareas, balance de carga, requerimientos de memoria, precisión de los cálculos, comunicaciones y sincronización entre procesos, errores de concurrencia, detección y tolerancia a fallos entre los más relevantes. La complejidad de los problemas requiere habilidades especiales de los desarrolladores: dominio de múltiples paradigmas de programación y frecuentemente múltiples lenguajes, conocimientos de redes y comprensión de la concurrencia y sus consecuencias. Por todo esto, se considera de gran interés el estudio de estos temas para el desarrollo de capacidades científico-tecnológicas en la UNdeC que favorezcan el trabajo interdisciplinario en la institución.

2. LÍNEAS DE I+D

Se abordaron las siguientes líneas de I+D desde la perspectiva del HPC como eje central:

- Análisis de la diversidad molecular de microorganismos del suelo [10]–[13]. Estudio e implementación de algoritmos que contribuyan a reducir los tiempos de procesamiento y aumentar la capacidad de análisis referidos a este campo de la bioinformática, a fin de profundizar en el estudio de la diversidad molecular de microorganismos del suelo asociados a cultivos regionales.
- Evaluación de enfoques de desarrollo HDL y HLL en FPGA para aplicaciones de procesamiento de imágenes [14]–[16]. Estudio de lenguajes de desarrollo HDL y HLL en FPGA para implementar soluciones de procesamiento de imágenes eficientes.
- Identificación biométrica masiva mediante venas del dedo usando redes de aprendizaje extremo (ELM) [17], [18]. Estudio de técnicas de computación paralela para mejorar la eficiencia y aceleración del preprocesamiento, extracción de características biométricas, búsqueda e identificación de individuos, y el diseño de algoritmos de ELM mejorados que manejen eficientemente lotes de datos de gran tamaño.

- Servicios basados en lingüística computacional para análisis de texto [19]–[24]. Estudio sobre modelos computacionales que reproduzcan aspectos del lenguaje humano, con el fin de realizar análisis lingüísticos como servicios para el Centro de Escritura en la UNdeC.
- Documentos inteligentes a través del Blockchain [25]–[27]. Estudio de la tecnología blockchain para garantizar la integridad de documentos universitarios.
- Nodo de información meteorológica [28], [29]. Estudio, diseño e implementación de algoritmos para reducir los tiempos de procesamiento, aumentar la capacidad de análisis y favorecer la escalabilidad de aplicaciones de análisis y proyección de datos climáticos.
- Desarrollo de algoritmos de procesamiento paralelo que permitan optimizar el cálculo de la Transformada Rápida de Fourier y operaciones matriciales, utilizadas en la medición de la intensidad compleja del sonido y otras variables de energética acústica [30], [31].

3. RESULTADOS

Se alcanzaron los siguientes resultados:

- Tres publicaciones en revista [32]–[34] y una en revisión.
- Diez presentaciones en congresos y/o workshops [35]–[44].
- Siete charlas científicas con invitados externos expertos en los distintos temas del proyecto.
- Cinco capacitaciones en otros centros de formación del país y del exterior en temas relacionados con el proyecto.
- Desarrollo de un curso de posgrado sobre “Análisis de rendimiento de aplicaciones paralelas (colaboración UNCo / UNdeC)” y una capacitación sobre “Eficiencia Energética en Sistemas de Cómputo” (colaboración UNCo / UNdeC / Red UNCI).
- Una tesis de maestría terminada (codirigida con UNLP), una tesis de maestría en etapa de defensa (codirigida con UCM-Chile), dos en desarrollo (una

codirigida con UNCo y otra codirigida con la UNSJ).

- Una tesina de grado en desarrollo (Asistente virtual académico de estudiantes de IS de la UNdeC).
- Un documento de recomendaciones para escribir tesinas de Ingeniería en Sistemas (en desarrollo).
- Un becario CIN.
- Tres trabajos finales de asignatura en los temas “Comparación entre Tecnologías de Programación Heterogénea CUDA y OpenCL” y “Cluster Beowulf con CUDA, OpenMP y OpenMPI: Reporte Técnico”
- Puesta en funcionamiento del cluster del Laboratorio de Sistemas Paralelos.
- Puesta en funcionamiento del servidor de altas prestaciones recientemente adquirido con fondos del Plan de Mejoramiento de la función de I+D+i del MINCyT (ver contexto).

4. FORMACIÓN DE RECURSOS HUMANOS

Seis miembros del equipo poseen formación de postgrado a nivel de doctorado, uno de ellos es especialista en Cómputo de Altas Prestaciones. Uno completó su formación a nivel de maestría durante 2021, otro se encuentra en la etapa de defensa y otro en proceso de desarrollo en temas abordados por esta propuesta. Las tres tesis de maestría son codirigidas por docentes de la Universidad Católica de Maule (Chile), la Universidad Nacional de La Plata y la Universidad Nacional del Comahue. Nueve miembros están cursando la Especialización en inteligencia de datos orientada a Big Data (UNLP), ya en etapa de elaboración de sus trabajos finales. Cada línea I+D propuesta integra al menos un docente investigador experto en el campo de cada estudio específico. Todos los temas propuestos se trabajan con estudiantes de grado de las carreras Ingeniería en Sistemas y Licenciatura en Sistemas de la UNdeC (ambas acreditadas por CONEAU). Los docentes forman parte de los equipos de diversas asignaturas de estas carreras, entre las que se encuentran programación, arquitecturas

de computadoras y arquitecturas paralelas. Nueve docentes se encuentran categorizados en el programa de incentivos.

5. BIBLIOGRAFÍA

- [1] S. Borkar y A. A. Chien, «The Future of Microprocessors», *Commun ACM*, vol. 54, n.º 5, pp. 67–77, may 2011, doi: 10.1145/1941487.1941507.
- [2] K. Ahmed y K. Schuegraf, «Transistor wars», *IEEE Spectr.*, vol. 48, n.º 11, pp. 50–66, nov. 2011, doi: 10.1109/MSPEC.2011.6056626.
- [3] V. V. Kindratenko *et al.*, «GPU clusters for high-performance computing», en *2009 IEEE International Conference on Cluster Computing and Workshops*, 2009, pp. 1–8.
- [4] J. Jeffers, J. Reinders, y A. Sodani, *Intel Xeon Phi Processor High Performance Programming: Knights Landing Edition*. Morgan Kaufmann, 2016.
- [5] A. Grama, A. Gupta, G. Karypis, y V. Kumar, *Introduction to Parallel Computing - Second Edition*. Pearson Education and Addison Wesley, 2003.
- [6] G. Hager y G. Wellein, *Introduction to high performance computing for scientists and engineers*. Boca Raton, FL: CRC Press, 2011.
- [7] J. Dongarra *et al.*, *Sourcebook of parallel computing*, vol. 3003. Morgan Kaufmann Publishers San Francisco, 2003.
- [8] F. E. Frati, «Software para arquitecturas basadas en procesadores de múltiples núcleos», Tesis, Facultad de Informática, 2015.
- [9] S. Ghosh, *Distributed Systems: An Algorithmic Approach*, 1.ª ed. University of Iowa, Iowa City, USA: Chapman and Hall/CRC, 2006.
- [10] M. S. De, M. Prager, R. E. Naranjo, y O. E. Sanclemente, «El suelo, su metabolismo, ciclaje de nutrientes y prácticas agroecológicas», *Agroecología*, vol. 7, n.º 1, pp. 19–34, 2012.
- [11] J. P. Hulsenbeck y F. Ronquist, «MrBayes: Bayesian inference of phylogeny», *Bioinformatics*, vol. 17, pp. 754–755, 2001.
- [12] A. Stamatakis, «RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies», *Bioinformatics*, vol. 30, n.º 9, pp. 1312–1313, may 2014, doi: 10.1093/bioinformatics/btu033.
- [13] M. A. Suchard, P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, y A. Rambaut, «Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10», *Virus Evol.*, vol. 4, n.º 1, ene. 2018, doi: 10.1093/ve/vey016.
- [14] R. Nane *et al.*, «A Survey and Evaluation of FPGA High-Level Synthesis Tools», *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 35, n.º 10, pp. 1591–1604, oct. 2016, doi: 10.1109/TCAD.2015.2513673.
- [15] S. Windh *et al.*, «High-Level Language Tools for Reconfigurable Computing», *Proc. IEEE*, vol. 103, n.º 3, pp. 390–408, mar. 2015.
- [16] R. Tessier, K. Pocek, y A. DeHon, «Reconfigurable Computing Architectures», *Proc. IEEE*, vol. 103, n.º 3, pp. 332–354, mar. 2015.
- [17] D. Ezhilmaran y P. R. B. Joseph, «A STUDY OF FEATURE EXTRACTION TECHNIQUES AND IMAGE ENHANCEMENT ALGORITHMS FOR FINGER VEIN RECOGNITION», p. 8, 2015.

- [18] A. Akusok, K. Björk, Y. Miche, y A. Lendasse, «High-Performance Extreme Learning Machines: A Complete Toolbox for Big Data Applications», *IEEE Access*, vol. 3, pp. 1011-1025, 2015.
- [19] M. Vallez y R. Pedraza, «El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines», *Hipertext Net*, 2007.
- [20] J. Texier, F. E. Frati, F. B. Carmona, A. E. Riba, M. Pérez, y J. Zambrano, «La gestión de la información en abierto, vehículo importante para maximizar la visibilidad web», presentado en XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016, Entre Ríos, Argentina), may 2016.
- [21] P. Gamallo Otero, J. C. Pichel Campos, M. García González, J. M. Abuín Mosquera, y T. Fernández Pena, «Análisis morfosintáctico y clasificación de entidades nombradas en un entorno Big Data», 2014.
- [22] S. W. D. Chien, C. P. Sishtla, S. Markidis, J. Zhang, I. B. Peng, y E. Laure, «An Evaluation of the TensorFlow Programming Model for Solving Traditional HPC Problems», en *International Conference on Exascale Applications and Software*, 2018, p. 34.
- [23] H. Guan, X. Shen, y H. Krim, «Egeria: A Framework for Automatic Synthesis of HPC Advising Tools Through Multi-layered Natural Language Processing», en *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, New York, NY, USA, 2017, p. 10:1–10:14.
- [24] P. Suber, *Ensuring open access for publicly funded research*. British Medical Journal Publishing Group, 2012.
- [25] A. Preukschat, *Blockchain: la revolución industrial de internet*. Gestión 2000, 2017.
- [26] L.-Y. Yeh, P. J. Lu, y J.-W. Hu, «NCHC blockchain construction platform (NBCP): rapidly constructing blockchain nodes around Taiwan», en *Digital Libraries (JCDL), 2017 ACM/IEEE Joint Conference on*, 2017, pp. 1–2.
- [27] H. Dai *et al.*, «TrialChain: A Blockchain-Based Platform to Validate Data Integrity in Large, Biomedical Research Studies», *ArXiv Prepr. ArXiv180703662*, 2018.
- [28] A. Botta, W. De Donato, V. Persico, y A. Pescapé, «On the integration of cloud computing and internet of things», en *Future internet of things and cloud (FiCloud), 2014 international conference on*, 2014, pp. 23–30.
- [29] P. Yue, H. Zhou, J. Gong, y L. Hu, «Geoprocessing in cloud computing platforms—a comparative analysis», *Int. J. Digit. Earth*, vol. 6, n.º 4, pp. 404–425, 2013.
- [30] D. Stanzial y C. E. Graffigna, «On the general connection between wave impedance and complex sound intensity», New Orleans, Louisiana, 2017, p. 055013. doi: 10.1121/2.0000797.
- [31] D. Stanzial y C. E. Graffigna, «Precision device for measuring the three dimensional spectra of complex intensity», New Orleans, Louisiana, 2017, p. 055014. doi: 10.1121/2.0000798.
- [32] R. Hernández-García *et al.*, «Fast Finger Vein Recognition Based on Sparse Matching Algorithm under a Multicore Platform for Real-Time Individuals Identification», *Symmetry*, vol. 11, n.º 9, p. 1167, sep. 2019, doi: 10.3390/sym11091167.
- [33] S. Guidet, R. J. Barrientos, R. Hernández-García, y F. E. Frati, «Exhaustive similarity search on a many-core architecture for finger-vein massive identification», *J. Phys. Conf. Ser.*, vol. 1702, p. 012012, nov. 2020.
- [34] R. Millón, E. Frati, y E. Rucci, «A Comparative Study between HLS and HDL on SoC for Image Processing Applications», *Elektron*, vol. 4, n.º 2, pp. 100-106, dic. 2020.
- [35] S. Guidet y E. Frati, «Desarrollo de un método para identificación de personas por venas de dedo en grandes bases de datos.», presentado en VIII Jornadas Científicas de Estudiantes Investigadores, Chilecito, La Rioja, Argentina, oct. 2019.
- [36] A. Riperto y J. Texier, «Smart Contract en la generación de título de grado bajo infraestructura HPC», presentado en VII Jornadas Científicas de Estudiantes Investigadores, Chilecito, La Rioja, Argentina, oct. 2019.
- [37] A. Ortiz y I. Flores, «Sistema para el análisis de textos científicos a través de HPC», presentado en VIII Jornadas Científicas de Estudiantes Investigadores, Chilecito, La Rioja, Argentina, oct. 2019.
- [38] J. Zambrano y J. Texier, «Laboratorio de escritura: dispositivos de acompañamiento en escritura para estudiantes de la UNDeC.», presentado en 1er Congreso Internacional de Ingeniería Aplicada de Ibero-American Science & Technology Education Consortium, 2019.
- [39] R. Hernández-García, S. Guidet, R. J. Barrientos, y F. E. Frati, «Massive Finger-vein Identification based on Local Line Binary Pattern under Parallel and Distributed Systems», en *2019 38th International Conference of the Chilean Computer Science Society (SCCC)*, 2019, pp. 1–7.
- [40] R. Millon, E. Rucci, y E. Frati, «Análisis Comparativo de Implementaciones HLS de Filtro Sobel en SoC», en *Libro de actas del XXVI Congreso Argentino de Ciencias de la Computación - CACIC 2020*, Buenos Aires, Argentina, oct. 2020, pp. 639-648.
- [41] R. Millon, F. E. Frati, y E. Rucci, «Implementación de Filtro de Detección de Bordes Sobel en SoC usando Síntesis de Alto Nivel», presentado en Congreso Argentino de Sistemas Embebidos (CASE2020)
- [42] S. Guidet, R. J. Barrientos, F. E. Frati, y R. Hernández-García, «Finger-vein individuals identification on massive databases», presentado en VIII Conference on Cloud Computing, Big Data & Emerging Topics, 2020.
- [43] F. E. Frati *et al.*, «Software y aplicaciones en computación de altas prestaciones para el contexto de la UNDeC», presentado en XXII Workshop de Investigadores en Ciencias de la Computación (WICC 2020, El Calafate, Santa Cruz), 2020.
- [44] F. E. Frati *et al.*, «Software y aplicaciones en computación de altas prestaciones para el contexto de la UNDeC», presentado en XXI Workshop de Investigadores en Ciencias de la Computación (WICC 2019, Universidad Nacional de San Juan)., abr. 2019.

TOMA DE DECISIONES EN SISTEMAS DE EVENTOS MEDIANTE INTELIGENCIA COMPUTACIONAL Y COMPUTACIÓN DE ALTO DESEMPEÑO

Esteban Schab^(1,3), Mariela Lopresti⁽²⁾, Natalia Miranda⁽²⁾,
Carlos Casanova^(1,3) y Fabiana Piccoli^(1 y 2)

⁽¹⁾ Universidad Autónoma de Entre Ríos, Concepción del Uruguay

⁽²⁾ LIDIC- Univ. Nacional de San Luis, San Luis

⁽³⁾ Univ. Tecnológica Nacional, Facultad Regional Concepción del Uruguay, Concepción del Uruguay
Argentina

{omlopres,nemiran, mpiccoli}@unsl.edu.ar
{schabe, casanovac}@frcu.utn.edu.ar

RESUMEN

La toma de decisiones en contextos dominados por grandes volúmenes de datos debe conjugar dos atributos usualmente contrapuestos: calidad y velocidad. La disponibilidad de información generada por personas y dispositivos abre nuevos desafíos en el diseño de mecanismos para aprovecharla. Estos deben ser capaces de determinar las decisiones de mayor utilidad sujetas a ventanas temporales que garanticen su factibilidad. Uno de estos mecanismos lo constituyen las distintas analíticas, las cuales buscan transformar los datos en información a través de técnicas diversas. En este trabajo proponemos dos líneas de investigación, una enfocada en la analítica prescriptiva, capaz de determinar acciones a ser ejecutadas en el momento (decisiones operativas) o en el futuro (decisiones tácticas para corto y mediano plazo, decisiones estratégicas para largo plazo) para lograr un objetivo deseado; la otra relacionada con las técnicas de aprendizaje supervisado y la recuperación de información no estructurada. En ambos, se propone la utilización de desarrollos provenientes de la Inteligencia Computacional y de la Computación de Alto Desempeño con el fin de obtener, de forma colaborativa, calidad y velocidad en las decisiones.

Palabras clave: Inteligencia Computacional. Analíticas. Big Data. Computación de Alto Desempeño.

CONTEXTO

Esta propuesta de trabajo se lleva a cabo dentro de los proyectos de investigación: “Tecnologías Avanzadas aplicadas al Procesamiento de Datos Masivos” (LIDIC, UNSL) y “Cómputo de Altas Prestaciones

aplicado a la Solución de Grandes Problemas” (UADER).

1. INTRODUCCIÓN

La mejora continua y adaptativa de los procesos de negocio resulta clave para mantener la competitividad de las organizaciones. En este contexto, la digitalización de los procesos, así como el incremento en las tecnologías de monitoreo, han llevado a producir una enorme cantidad de datos, los cuales tienen un gran potencial para mejorar los procesos conducidos por analíticas [9][12][30].

Las analíticas buscan transformar los datos en conocimiento para la toma de decisiones [11], y se distinguen cuatro tipos de analíticas según el nivel de automatización del proceso [17]. Ellas son:

- Descriptivas: intentan responder qué ha pasado o está pasando.
- Diagnósticas: por qué ha pasado o está pasando, analizando para ello datos históricos [26].
- Predictivas: buscan responder qué sucederá, aplicando el conocimiento para predecir nuevos datos sobre el presente o el futuro (pronóstico) [6].
- Prescriptivas: responden qué debería hacerse para lograr un objetivo, esto es, determina acciones a ser ejecutadas en el momento (decisiones operativas) o en el futuro (decisiones tácticas para corto y mediano plazo, decisiones estratégicas para largo plazo).

Cabe aclarar que ninguno de los tres primeros enfoques sugiere acciones concretas, sino que descansan en el juicio subjetivo y las habilidades analíticas del usuario para deducir acciones de mejora [26] [27].

Si bien el volumen de datos generados por personas y dispositivos se encuentra en continuo crecimiento, y a pesar de los avances tecnológicos, en general, las analíticas de procesos existentes dentro de la industria actual no aprovechan completamente el conocimiento oculto debido a las siguientes limitaciones [9]:

- a. No hacen uso de técnicas prescriptivas para transformar los resultados del análisis en acciones de mejora concretas, dejando este paso completamente a criterio del usuario.
- b. Hacen un uso intensivo de datos de sistemas en producción, generando un deterioro en el desempeño de las herramientas de software que soportan los procesos.
- c. La optimización es conducida *ex post*, después de completado el proceso, en contraste a la mejora proactiva durante la ejecución del proceso.

En el área de datos masivos (o *Big Data*), se identifica como área emergente el procesamiento de *datastreams*, también llamado *Data Stream Mining* [2][14][24]. Un *datastream* es una representación digital y transmisión continua de datos, los cuales describen una clase de eventos relacionada [23] [32]. Mediante el procesamiento de estos *datastreams* se puede lograr la respuesta en tiempo real a los eventos en forma de toma de decisiones.

Los grandes volúmenes de datos generados pueden ser utilizados en algoritmos de diversa índole para la generación de analíticas, particularmente las descriptivas y predictivas, utilizando por ejemplo aprendizaje supervisado o no supervisado. Existe, sin embargo, una dificultad en lo relativo a las analíticas prescriptivas: no cuentan con un “profesor” [25] que les enseñe qué acción tomar en cada circunstancia. Un tipo de aprendizaje que no necesita de un profesor es el llamado aprendizaje por refuerzo. En este esquema es el propio agente quien es capaz de juzgar y criticar sus acciones con base en sus percepciones y de alguna medida de aptitud, recompensa o refuerzo. La tarea del aprendizaje por refuerzo es usar recompensas observadas para aprender una política óptima (o aproximadamente óptima) del entorno, sin asumir ningún conocimiento *a priori* [25][29]. Esta política le dice al agente qué hacer en cada estado posible a alcanzar.

Por otra parte, la Inteligencia Computacional (IC), que toma su inspiración de la naturaleza [13], apunta a resolver los problemas aprovechando la imprecisión y la incertidumbre presente en el proceso de toma de decisiones brindando las “soluciones más satisfactorias”. Por ello, el aprendizaje por refuerzo en combinación con técnicas de IC son un activo fundamental para la elaboración de analíticas prescriptivas que permitan resolver problemas concernientes a la toma de decisiones en contextos de incertidumbre.

Cualquiera sea la analítica, su generación debe ser lo suficientemente rápida como para procesar los *datastreams* generados en forma continua por el sistema, esto implica el uso de técnicas y herramientas de Computación de Alto Desempeño (HPC).

En la siguiente sección se describen con mayor detalle las características de las líneas de investigación.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

En este contexto de Procesos de negocio y su mejora, los *datastreams* abren nuevas y amplias oportunidades para la creación de valor en las organizaciones. A través de su procesamiento se puede conseguir respuesta en tiempo real a eventos en forma de toma de decisiones.

En [8][10][32] encontramos distintos enfoques para la definición de eventos en un proceso de negocio, variando en su complejidad. Una secuencia o flujo de eventos (*Event Stream*) es una serie ordenada y potencialmente ilimitada de eventos [21]. Los flujos de eventos se generan y utilizan en muchos sistemas. Ejemplo de ellos son:

- Bancos y otros sistemas de atención al público: En ellos la generación de *datastreams* es causada por los sistemas de gestión de atención implementados [19]. Estos sistemas en general utilizan modelos relacionales de bases de datos y, si bien permiten la elaboración de analíticas, puede resultar inapropiada su implementación. En un contexto de *Software as a Service* (SaaS) [31], el desempeño del sistema de atención, puede colapsar ante las continuas consultas

por monitoreo. Por tal motivo, la generación y el procesamiento en paralelo al sistema de atención de los eventos como *datastreams*, puede ser la tecnología de base para un monitoreo eficiente.

- Enrutamiento de vehículos: conocido por sus siglas VRP (*Vehicle Routing Problem*) [3], con suministro de información y reencaminamiento en tiempo real, orientado a la búsqueda de un paradigma de movilidad inteligente [16]. Dentro de este problema se pueden estudiar de forma particular o en conjunto la logística urbana, el transporte de personas y los conductores individuales. En este caso los *datastreams* son generados de forma distribuida por cada agente involucrado y pueden ser procesados de forma centralizada o distribuida según el esquema elegido y los recursos disponibles.

Partiendo del objetivo general de lograr la Optimización de procesos de negocio por recomendación, surgen dos líneas de investigación. La primera propone la composición de modelos de analítica prescriptiva para superar los inconvenientes descritos. Estos modelos serán parte esencial de un proceso de mejora continua basado en la recomendación de acciones operativas y tácticas a fin de mantener el rendimiento del sistema en los valores deseados.

La segunda línea, está relacionada con el aprendizaje supervisado. Los algoritmos de aprendizaje automático pueden resolver de manera eficiente problemas con grandes conjuntos de datos complejos de distintas fuentes: web, redes sociales, eventos, teléfonos, telescopios, imágenes satelitales, entre otros. El algoritmo k-NN es muy conocido y ampliamente utilizado por su sencillez y robustez, descrito en 1967 por Cover y Hart [5]. Este método se basa en la idea intuitiva de que objetos similares pertenecen a la misma clase, entonces la clase de un objeto puede ser inferida a partir de la clase a la que pertenecen los objetos (o el objeto) más parecidos de la muestra de aprendizaje. La idea de similitud se refleja formalmente en el concepto de distancia.

La implementación más simple de k-NN tiene como gran inconveniente determinar los k-vecinos más cercanos de un objeto. Para esto se

requiere calcular las distancias entre él y todos los demás objetos de entrenamiento. Con el tamaño creciente de los conjuntos de datos, este enfoque se vuelve ineficiente, resultando en una búsqueda de complejidad lineal. Se han desarrollado variantes de k-NN, las cuales no sólo consideran la forma en que se comparan los objetos, sino también en las técnicas de programación, como es el caso de HPC, en particular GPU.

Para la primera línea, se propone el uso de agentes de aprendizaje por refuerzo [29], junto a técnicas provenientes de la Inteligencia Computacional: redes neuronales como modelos, teoría de conjuntos difusos como lenguaje de especificación, y métodos numéricos y metaheurísticos para el entrenamiento de tales modelos [7][28][33]. Para la segunda, el objetivo es acelerar y mejorar los métodos k-NN desarrollados considerando su utilización en problemas con grandes volúmenes de datos.

En ambos casos, se considera satisfacer la necesidad de dar rápida respuesta a los procesos de negocios dinámicos, en consecuencia es mandatorio pensar en la aplicación de modelos/paradigmas de HPC [9][15][22] portables, particularmente en GPU.

3. RESULTADOS OBTENIDOS/ESPERADOS

Por un lado, se encuentra en desarrollo un agente de aprendizaje por refuerzo para un caso de enrutamiento de vehículos, con suministro de información y re-encaminamiento en tiempo real. Este caso es uno de logística urbana. Se espera completar el desarrollo de un modelo prescriptivo dirigido por los datos, basado en la recomendación automática y proactiva de acciones operativas y tácticas destinadas a mantener los indicadores de rendimiento del sistema dentro de los valores deseados. Por el otro, se está trabajando en la adecuación de las implementaciones paralelas de k-NN considerando múltiples GPU.

Como uno de los objetivos es lograr soluciones paralelas portables, de costo predecible, capaz de explotar las ventajas de modernos ambientes HPC a través de

herramientas y “frameworks de computación” de alto nivel [15][22], los primeros desarrollos se están haciendo en GPU Nvidia con CUDA [18], y se prevé el uso de otras tecnologías como OpenCL[4] y OneAPI[20].

4. FORMACIÓN DE RECURSOS HUMANOS

Los resultados esperados respecto a la formación de recursos humanos son el desarrollo de 1 tesis de doctorado, 2 de maestría y de varias tesinas de grado en las universidades intervinientes.

5. BIBLIOGRAFÍA

- [1] M.Barrionuevo, M.Lopresti, N.Miranda, and M.F. Piccoli. “Solving a big-data problem with gpu: the network traffic analysis”. *Journal of Comp. Sc. and Techn.*, 15(01): Pp.30–39, Apr. 2015.
- [2] A. Bifet and J. Read. “Ubiquitous artificial intelligence and dynamic data streams”. In *Proceedings of the 12th ACM International Conference on Distributed and Event-Based Systems, DEBS '18*, Pp. 1–6, Association for Computing Machinery. New York, USA, 2018.
- [3] K. Braekers, K. Ramaekers and I. Van Nieuwenhuysse, "The vehicle routing problem: State of the art classification and review", *Comp. & Ind. Eng.*, vol. 99, pp. 300-313, 2016.
- [4] M. Breyer, G. Daiss and D. Pflüger. “*Performance-Portable Distributed k-Nearest Neighbors Using Locality-Sensitive Hashing and SYCL*”. *Int. Workshop on OpenCL*. ISBN 9781450390330. Association for Computing Machinery. 2021
- [5] T. Cover and P. Hart. 1967. *Nearest neighbor pattern classification*. *IEEE Transactions on Information Theory* 13, 1, 7. Jan. 1967. doi.org/10.1109/tit.1967.1053964.
- [6] G. Deka. "Big data predictive and prescriptive analytics". In *Handbook of research on cloud infrastructures for Big Data analytics*. IGI Global. p. 370-391. 2014.
- [7] A. Ebrahimnejad and J. L. Verdegay. “Fuzzy sets-based methods and techniques for modern analytics”. Springer Int. Publishing. 2018.
- [8] O. Etzion and P. Niblett, "Event processing in action", Manning, 2011.
- [9] C. Gröger, H. Schwarz, and B. Mitschang. “Prescriptive analytics for recommendation-based business process optimization”. In *Int. Conf. on Business Information Systems*, pages 25–37. Springer, 2014.
- [10] A. Hinze, K. Sachs and A. Buchmann. "Event-based applications and enabling technologies". *Proc. Third ACM Int. Conf. on Dist. Event-Based Systems*. pp. 1-15. 2009.
- [11] C. Holsapple, A. Lee-Post and R. Pakath. “A unified foundation for business analytics. *Decision Support Systems*”, 64:130{141, 2014.
- [12] M. Kaur Saggi and S. Jain. “A survey towards an integration of big data analytics to big insights for value-creation”. *Inf. Proc.& Management*, 54(5):758-790, 2018.
- [13] J. Keller, D. Liu and D. Fogel. "Fundamentals of computational intelligence: neural networks, fuzzy systems, and evolutionary computation". Wiley & Sons. 2016.
- [14] T. Kolajo, D. Olawande and A. Ayodele. "Big data stream analysis: a systematic literature review". *Journal of Big Data*, vol. 6, no 1, pp. 1-30, 2019.
- [15] S. Kurgalin and S. Borzunov, “A Practical Approach to High-Performance Computing”. Springer. 2019.
- [16] S. Melo, J. Macedo and P. Baptista, "Guiding cities to pursue a smart mobility paradigm: An example from vehicle routing guidance and its traffic and operational effects". *Research in Transp. Economics*, vol. 65, p. 24-33, 2017.
- [17] M. Minelli, M. Chambers, and A. Dhiraj. “Big data, big analytics: emerging business intelligence and analytic trends for today’s businesses”. Vol. 578. Wiley & Sons, 2013.
- [18] Nvidia. “*CUDA C++ Programming Guide, Design Guide*”.

- https://docs.nvidia.com/cuda/pdf/CUDA_C_Programming_Guide.pdf. 2019.
- [19] T. Olanrewaju. "The rise of the digital "bank."<https://www.mckinsey.com/business-functions/digital-mckinsey/ourinsights/the-rise-of-the-digital-bank/>. Accessed: 2020.
- [20] OneAPI Specification, release 1.0-rev 3. <https://spec.oneapi.com/versions/1.0-rev-3/>. 2020.
- [21] T. J. Owens, "Survey of event processing", Air Force Research Lab Rome NY Inf. Directorate, 2007.
- [22] P. Pacheco. "An Introduction to Parallel Programming", 1Ed., San Francisco, USA: Morgan Kaufmann Publishers Inc., 2011.
- [23] F. Pigni, G. Piccoli, and R. Watson. "Digital data streams: Creating value from the real-time flow of big data". *California Management Review*, 58(3):5–25, 2016.
- [24] S. Ramírez-Gallego, B. Krawczyk, S. García, M. Wozniak and F. Herrera. "A survey on data preprocessing for data stream mining: Current status and future directions". *Neurocomputing*, 239:39 – 57, 2017.
- [25] S. Russell and P. Norvig. "Inteligencia Artificial: un enfoque moderno". 2004.
- [26] R. Sharda, D. Delen, and E. Turban. "Business Intelligence and Analytics". 2015.
- [27] R. Sharda, D. Delen, and E. Turban. "Analytics, Data Science, & Artificial Intelligence". 2020.
- [28] N. Siddique and H. Adeli. "Computational intelligence: synergies of fuzzy logic, neural networks and evolutionary computing". Wiley & Sons, 2013.
- [29] R. Sutton and A. Barto. "Reinforcement learning: An introduction". MIT press. 2018.
- [30] U. Thirathon, B. Wieder, Z. Matolcsy and M.L. Ossimitz. "Impact of big data analytics on decision making and performance". In *Int. Conf. on Enterprise Systems, Accounting and Logistics*, 2017.
- [31] M. Turner, D. Budgen and P. Brereton, "Turning software into a service". In *Computer*, vol. 36, no. 10, pp. 38-44, Oct. 2003.
- [32] C. Wrench, et al. "Data stream mining of event and complex event streams: A survey of existing and future technologies and applications in big data." *Enterprise Big Data Engineering, Analytics, and Management*. IGI Global, 2016. Pp. 24-47.
- [33] L.A. Zadeh. "Fuzzy logic, neural networks, and soft computing". *Comm. ACM* 37,3, 77–84. <https://doi.org/10.1145/175247.175255>. March 1994.

PSST – Procesamiento de Señales y Sistemas en Tiempo Real

Métricas de rendimiento, aplicaciones de datos masivos e inteligencia artificial en cómputo paralelo y distribuido

Javier Balladini¹, Marina Morán¹, Claudio Zanellato¹, Claudia Rozas¹, Rodrigo Cañibano¹,
Cristina Orlandi², Armando De Giusti³, Remo Suppi⁴, Dolores Rexachs⁴, Emilio Luque⁴,
Emmanuel Frati⁵

¹ Facultad de Informática, Universidad Nacional del Comahue
{javier.balladini, marina, claudio.zanellato, claudia.rozas, rcanibano}@fi.uncoma.edu.ar

² Hospital Francisco Lopez Lima - orlandi.mariacristina@gmail.com

³ Instituto de Investigación en Informática LIDI, Universidad Nacional de La Plata - degiusti@lidi.info.unlp.edu.ar

⁴ Departamento de Arquitectura de Computadores y Sistemas Operativos, Universidad Autónoma de Barcelona
{remo.suppi, dolores.rexachs, emilio.luque}@uab.es

⁵ Universidad Nacional de Chilecito, fefrati@undec.edu.ar

Resumen

Los grandes avances tecnológicos de los sistemas de cómputo paralelo y distribuido hacen viable nuevas soluciones a problemas. Por un lado, nos enfocamos en métricas de rendimiento. En particular, nos abocamos a métricas energéticas que cobraron enorme relevancia debido al gran número de unidades de procesamiento que componen los sistemas de cómputo. Por otro lado, buscamos aplicar técnicas de cómputo paralelo y distribuido para brindar soluciones en el sector salud. En especial, nos orientamos a sistemas de alertas tempranas de gravedad, basadas en inteligencia artificial. Una aplicación está destinada a Unidades de Cuidados Intensivos, que debe tratar con datos masivos, y otra aplicación tiene como fin la clasificación del nivel de gravedad de pacientes con COVID-19, que presenta una arquitectura distribuida, tolerante a fallos y de simple administración. Los trabajos se desarrollan en colaboración con otras universidades, y un hospital público de Argentina. La formación de recursos humanos en estas líneas está orientada al nivel de tecnicatura, grado, maestría y doctoral.

Palabras claves: computación de altas prestaciones, eficiencia energética, big data, salud, inteligencia artificial.

1. Contexto

Las líneas de investigación aquí presentadas están enmarcadas dentro del proyecto de investigación 04/F017 "Cómputo paralelo y distribuido: métricas de rendimiento, aplicaciones de big data e inteligencia artificial", financiado por la Universidad Nacional del Comahue (UNComa), con inicio el 01/01/2021 y finalización el 31/12/2024, acreditado por el Ministerio de Educación de Argentina.

La mayoría de los temas de investigación surgieron en proyectos anteriores acreditados. El eje de aplicaciones para la salud, que involucra temáticas de inteligencia artificial y Big Data, se desarrolla en colaboración con el Hospital Francisco López Lima de General Roca (Río Negro) y con interés del Ministerio de Salud de la Provincia de Río Negro. El eje de métricas de rendimiento de los sistemas de cómputo paralelo, se desarrolla en colaboración con el Instituto de Investigación en Informática LIDI de la Universidad Nacional de La Plata, el grupo de investigación "High Performance Computing for Efficient Applications and Simulation" de la Universidad Autónoma de Barcelona, España, y con la Universidad Nacional de Chilecito, La Rioja.

2. Introducción

Numerosas aplicaciones informáticas que requieren el procesamiento de grandes volúmenes de datos están surgiendo gracias al uso de tecnologías de cómputo paralelo y distribuido, que hasta hace unos años eran excepcionales. La masificación tecnológica y consecuente reducción de costos hizo que se pueda pensar en el desarrollo de nuevas aplicaciones que tengan un fuerte impacto social. En particular, nos interesamos principalmente en el sector de salud.

A su vez, nos enfocamos en las métricas de eficiencia computacional y energética de los sistemas de cómputo paralelo y distribuido. Desde los inicios de la computación las métricas de interés se relacionaban únicamente con la eficiencia computacional. Actualmente, las métricas de rendimiento energéticas en muchos casos superan en relevancia a la velocidad de cómputo. En particular, nos hemos orientado a la gestión del consumo energético de estos sistemas, para reducir su impacto económico, medioambiental y social.

A continuación se introduce la motivación y problemática de ambas líneas de investigación.

2.1. Aplicaciones para la salud

Unidad de Cuidados Intensivos

Una Unidad de Cuidados Intensivos (UCI) provee cuidados continuos y rigurosos a personas adultas críticamente enfermas que pueden beneficiarse de tratamiento, y da un buen morir a pacientes irrecuperables. Los datos de los pacientes involucran datos clínicos de baja frecuencia y flujos de datos fisiológicos de alta frecuencia generados por el equipamiento médico. En una UCI típica, los enfermeros completan manualmente datos en formularios, registrando datos clínicos y fisiológicos. Los datos fisiológicos se obtienen por observación de las pantallas del equipamiento médico a intervalos de tiempo que definen los médicos para cada paciente. El equipamiento médico emite alertas cuando hay riesgo en la salud del paciente basándose en mediciones de parámetros. Luego, los médicos analizan los datos de los formularios y dan

a los enfermeros indicaciones de tratamientos.

Los principales problemas que ocurren en las UCIs típicas son: (a) gestión de la información es proclive a errores humanos, (b) pérdida de datos entre registros de enfermería, (c) detección tardía del deterioro de la salud de los pacientes. Las causas tienen origen en el tratamiento manual de la información (lento y propenso a errores), y cuyos efectos incluyen diagnósticos imprecisos o incorrectos o retrasados, inconsistencias en la información, menos datos para investigaciones médicas, y mayores necesidades de recursos humanos. En la mayoría de las UCIs del país, no hay médicos intensivistas durante las 24hs, y hay sobrecarga laboral, impactando negativamente en la salud de los pacientes.

Las soluciones se orientan a sistema expertos de detección automática, temprana y progresiva del deterioro de pacientes, tal como los propuestos en [1, 10, 13].

DetECCIÓN DE RIESGO EN PACIENTES CON COVID-19

El 11 de marzo de 2020 la Organización Mundial de la Salud dictaminó la pandemia de COVID-19. El sistema sanitario argentino tiene su límite crítico en la cantidad de enfermeros disponibles por habitantes, con un número de 2,6 enfermos cada mil habitantes en contra parte de los 5,7 que tienen España e Italia¹, quienes no han podido contener la situación con ese personal. Esta situación generó que los recursos humanos especializados escaseen.

El “triage” es un método que permite la organización y optimización de los recursos en situaciones críticas, seleccionando y clasificando a los pacientes en diferentes niveles de gravedad para una correcta asignación de la atención. Este método fue aplicado en China [11] en pacientes con COVID-19 por medio de la adaptación de uno de los sistemas de alertas tempranas más conocido y validado. Esto permite mejorar el rendimiento del personal, reduciendo los controles en pacientes estables y

¹Datos relevados en 2017 por el Banco Mundial (<https://datos.bancomundial.org>).

aumentando los mismos en pacientes graves, y reducir la mortalidad hospitalaria inesperada en áreas de internación general.

Si bien el triage puede ser realizado mediante cálculos hechos a mano, no solo se agregaría una nueva tarea al personal de salud sino que sería un método muy propenso a errores humanos y por ende poco fiable. En consecuencia, estamos desarrollando un sistema informático inteligente que implemente un sistema de alerta temprana para pacientes en internación general con COVID-19, que realiza un triage a través del análisis automático y en tiempo real de datos clínicos, clasificando a los pacientes según el riesgo, emitiendo alertas al personal de salud, y permitiendo organizar y optimizar recursos humanos. El sistema, a partir de la carga de datos de enfermería, comorbilidades, resultados radiológicos y de laboratorio, predice la gravedad de los pacientes (bajo, moderado, alto, crítico). No conocemos actualmente sistemas de este tipo.

2.2. Métricas de rendimiento - eficiencia energética en HPC

La computación de alto rendimiento (HPC) sigue aumentando su rendimiento computacional y su eficiencia energética. Por ejemplo, el superordenador Fugaku, la supercomputadora más rápida del mundo en el ranking Top500, presenta 442 PFlops frente a los 148,6 PFlops de su predecesor, Summit. Estas máquinas, al momento de ingresar al Top500 se encontraban entre las diez de mayor eficiencia energética del ranking Green500, con alrededor de 15 GFlops/W. Sin embargo, como las máquinas aumentan enormemente de tamaño, el consumo de energía sigue incrementándose; mientras que la computadora Summit tiene un consumo de 10 MW, Fugaku tiene casi 30 MW (aproximadamente lo mismo que consume una ciudad de 400.000 habitantes). Como este aumento del consumo de energía no es sostenible, es necesario reducirlo.

La computación ecológica es el estudio y la práctica de la computación ambientalmente sostenible. Ella se ocupa de diferentes aspectos

de los sistemas de cómputo: diseño, manufactura, eliminación, y uso. Este último aspecto, el uso ecológico, se refiere al uso de los sistemas de cómputo con conciencia ambiental. Es posible reducir el consumo de energía de los sistemas de cómputo utilizando diferentes estrategias que deben ser consideradas a nivel del software [9, 6], y consisten en realizar cambios en la configuración del sistema (escalado dinámico de frecuencia y tensión, e hibernación de recursos) o en las aplicaciones (uso adecuado de la jerarquía de memoria, rediseño de algoritmos, planificación de tareas, y asignación de tareas a recursos hardware).

El aumento significativo de la cantidad de unidades de procesamiento causa el incremento del consumo de energía y la disminución de la confiabilidad del sistema de cómputo. Así, a poco tiempo de arribar a la era exaescala (o ya en ella para aplicaciones que requieren ejecutar operaciones de baja complejidad), la tolerancia a fallos y el consumo energético se han identificado como los dos mayores desafíos a enfrentar [12, 8]. En consecuencia, proponemos desarrollar metodologías, modelos y software para gestionar el consumo energético, en especial al utilizar mecanismos de tolerancia a fallos en máquinas paralelas de tipo cluster.

3. Líneas de investigación

El eje central de nuestra investigación es desarrollar metodologías, modelos y soluciones informáticas para colaborar en la resolución de problemas de cómputo paralelo y distribuido, que puedan tener una alta demanda computacional e impacto social en el campo de la salud, y en la reducción del consumo energético de sistemas de HPC.

3.1. Aplicaciones para la salud

UCI: Nuestro sistema intenta emular el comportamiento de un médico intensivista experto, dando recomendaciones para la toma de decisiones clínicas, con el objetivo de reducir la incertidumbre sobre el diagnóstico, las opciones de tratamiento y el pronóstico. La solución

requiere la aplicación de técnicas de cómputo paralelo y distribuido para procesamiento en tiempo real de algoritmos de inteligencia artificial sobre grandes volúmenes de datos. A diferencia de otras alternativas, nuestro objetivo se orienta a la construcción de un sistema multihospitalario (con el fin de incrementar el volumen de datos y consecuente extracción de conocimiento) con soporte de telemedicina, que integre componentes de software libre maduros, que nosotros optimizados para nuestro dominio, que sea seguro, tolerante a fallos y resiliente.

COVID-19: Nuestro objetivo es construir un sistema y modelo computacional de alerta temprana para pacientes con COVID-19, inicialmente basado en el conocimiento de personal experto en salud, y posteriormente mejorado mediante técnicas de aprendizaje automático. La complejidad de la arquitectura del sistema no está en la alta demanda computacional sino en evitar el uso de servidores estándares en los hospitales (a fin de simplificar la administración multihospitalaria), y que el sistema local de un hospital se mantenga operativo ante fallos de conexión con la nube, utilizando únicamente dispositivos móviles.

3.2. Métricas de rendimiento - consumo energético en HPC

Nos centramos en el desarrollo de metodologías, modelos y construcción de software para administrar y gestionar el consumo de energía y prestaciones computacionales de sistemas de cómputo paralelo. Actualmente, nuestro principal objetivo es la gestión energética en mecanismos de tolerancia a fallos basados en checkpoints.

4. Resultados y objetivos

4.1. Aplicaciones para la salud

UCI: En [5] presentamos un análisis del estado general de la UCI del hospital Francisco López Lima, y la propuesta del diseño de alto nivel del sistema. En [2] presentamos el diseño

de la infraestructura del sistema de procesamiento de reglas clínicas y un prototipo. En [3] propusimos la herramienta como mejora de la prestación de servicios integrados de cuidados intensivos de la salud. En [7] presentamos una optimización de la administración de datos de curvas fisiológicas. Próximos objetivos: continuar con el desarrollo de hardware y software para extraer datos del equipamiento médico, desarrollar aplicaciones para el procesamiento eficiente de señales, manejo de errores en datos fisiológicos, optimización de la infraestructura de datos masivos de tiempo real para el procesamiento de reglas clínicas, y el desarrollo de una aplicación para la interacción del sistema con médicos y enfermeros.

COVID-19: Estamos completando la primera versión del sistema informático (COVINDEX), que implementa modelos de alertas tempranas. El sistema se compone de una aplicación web, una aplicación móvil y una aplicación que ejecuta en un servidor en la nube. A partir de los datos históricos que se recolecten de pacientes, se calibrará periódicamente el sistema de alerta temprana de insuficiencia respiratoria, sobre la base de técnicas de aprendizaje automático, para aumentar la capacidad predictiva del sistema. Aún no se cuenta con publicaciones científicas. Próximos objetivos: estamos trabajando en conjunto con el Ministerio de Salud de la Provincia de Río Negro para construir un sistema adaptable a otras enfermedades y para el uso en pacientes con enfermedades no contempladas de manera específica, que permita derivaciones de pacientes entre hospitales y un control general de la situación por parte del ministerio.

4.2. Métricas de rendimiento - consumo energético en HPC

Los métodos de tolerancia a fallos tienen fuerte incidencia en el consumo energético de los sistemas de HPC, y resulta de suma importancia conocer, antes de ejecutar una cierta aplicación, el impacto que pueden producir los diferentes métodos y configuraciones del mismo. En [4], presentamos una metodología para

predecir el consumo energético producido por el método de checkpoint coordinado remoto, y en [15] expusimos un análisis de los factores que afectan el consumo energético de operaciones de checkpoint y restart en clusters. En [14] propusimos un modelo para estimar el consumo energético de operaciones de checkpoint y restart, y un método para su construcción. En [16] presentamos estrategias para checkpoints no coordinados que, al momento de un fallo de un nodo, permiten gestionar y reducir el consumo energético de los nodos que no han fallado; se construyó también un modelo energético y un simulador que permite evaluar las estrategias. Próximos objetivos: se está ampliando el simulador para contemplar más situaciones como las operaciones no bloqueantes y las esperas en cascada causadas por las dependencias entre procesos.

Se dictó un curso de posgrado “Análisis de rendimiento de aplicaciones paralelas” (colaboración UNCo / UNdeC) y una capacitación sobre “Eficiencia Energética en Sistemas de Cómputo” (colaboración UNCo / UNdeC / Red UNCI).

5. Formación de recursos humanos

El equipo de trabajo de la Universidad Nacional del Comahue tiene un Doctor y un Máster, una estudiante de Doctorado y dos estudiantes de Maestría. Tres estudiantes de grado están realizando trabajos de tesis, y hay un estudiante de tecnicatura. En 2021 no se ha finalizado ninguna tesis.

Referencias

- [1] S. Balaji, M. Patil, and C. McGregor. A cloud based big data based online health analytics for rural nicus and picus in india: Opportunities and challenges. In *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 385–390, 2017.
- [2] Javier Ballardini, Pablo Bruno, Rafael Zurita, and Cristina Orlandi. An automatic and early detection of the deterioration of patients in intensive and intermediate care units. *Journal of Computer Science and Technology*, 18(03):e25, Dec. 2018.
- [3] Javier Ballardini, Pablo Bruno, Rafael Zurita, Cristina Orlandi, Remo Suppi, Dolores Rexachs, and Emilio Luque. A tool for improving the delivery of integrated intensive health care performance. *International Journal of Integrated Care*, 19(4), 07/2019 2019.
- [4] Javier Ballardini, Marina Morán, Dolores Rexachs, and Emilio Luque. Metodología para predecir el consumo energético de checkpoints en sistemas de hpc. *XX Congreso Argentino de Ciencias de la Computación (CACIC 2014)*, 2014.
- [5] Javier Ballardini, Claudia Rozas, Emmanuel Frati, Nestor Vicente, and Cristina Orlandi. Big data analytics in intensive care units: challenges and applicability in an argentinian hospital. *Computer Science and Technology (JCST)*, 2015.
- [6] Aurelien Bouteiller, Franck Cappello, Jack Dongarra, Amine Guermouche, Thomas Héroult, and Yves Robert. Multi-criteria checkpointing strategies: Response-time versus resource utilization. In *European Conference on Parallel Processing*, pages 420–431. Springer, 2013.
- [7] Rodrigo Cañibano, Claudia Rozas, Cristina Orlandi, and Javier Ballardini. Data management optimization in a real-time big data analysis system for intensive care. In *Conference on Cloud Computing, Big Data & Emerging Topics*, pages 93–107. Springer, Cham, 2020.
- [8] Franck Cappello, Al Geist, William Gropp, Sanjay Kale, Bill Kramer, and Marc Snir. Toward exascale resilience: 2014 update. *Supercomputing Frontiers and Innovations*, 1(1), 2014.
- [9] Mohammed El Mehdi Diouri, Olivier Glück, Laurent Lefèvre, and Franck Cappello. Ecofit: A framework to estimate energy consumption of fault tolerance protocols for hpc applications. In *Proceedings of the 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing*, pages 522–529, 2013.
- [10] Jagreet Kaur and Dr. Kulwinder Singh Mann. AI based HealthCare platform for real time, predictive and prescriptive analytics using reactive programming. *Journal of Physics: Conference Series*, 933:012010, jan 2018.
- [11] Xuelian Liao, Bo Wang, and Yan Kang. Novel coronavirus infection during the 2019–2020 epidemic: preparing intensive care units-the experience in sichuan province, china. *Intensive care medicine*, 46(2):357–360, 2020.
- [12] Yongpeng Liu and Hong Zhu. A survey of the research on power management techniques for high-performance systems. *Software: Practice and Experience*, 40(11):943–964, 2010.
- [13] Fernando López-Martínez, Edward Rolando Núñez-Valdez, Vicente García-Díaz, and Zoran Bursac. A case study for a big data and machine learning platform to improve medical decision support in population health management. *Algorithms*, 13(4), 2020.
- [14] M. Morán, J. Ballardini, D. Rexachs, and E. Luque. Prediction of energy consumption by checkpoint/restart in hpc. *IEEE Access*, 7:71791–71803, 2019.
- [15] Marina Morán, Javier Ballardini, Dolores Rexachs, and Emilio Luque. Factores que afectan el consumo energético de operaciones de checkpoint y restart en clusters. *XXIV Congreso Argentino de Ciencias de la Computación (CACIC 2018)*, 2018.
- [16] Marina Morán, Javier Ballardini, Dolores Rexachs, and Enzo Rucci. Towards management of energy consumption in hpc systems with fault tolerance. In *2020 IEEE Congreso Biental de Argentina (ARGENCON)*, pages 1–8, 2020.

Modelos Matemáticos y Aritmética Computacional

Javier Giacomantone - Oscar Bria

Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – UNLP

Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)
La Plata, Buenos Aires

{jog, onb}@lidi.info.unlp.edu.ar

Resumen

Los modelos matemáticos determinísticos y probabilísticos permiten el análisis, diseño e implementación de distinto tipo de sistema. Los métodos computacionales y el desarrollo de técnicas aritméticas discretas en hardware y de algoritmos específicos nos permiten abordar problemas y plantear soluciones numéricas viables. El objetivo principal de las líneas de investigación y desarrollo descriptas es estudiar y evaluar modelos matemáticos, métodos numéricos y la implementación aritmética computacional que permita abordar problemas específicos en ingeniería. Los modelos abordados son dependientes del tipo de sistema estudiado y del fenómeno analizado. Determinar el método para evaluar el rendimiento y las soluciones numéricas óptimas o sub-óptimas forma parte de los objetivos generales.

Palabras Clave: modelos matemáticos aplicados, métodos computacionales, aritmética computacional, desempeño de sistemas, confiabilidad de sistemas.

Contexto

Las líneas de investigación y desarrollo (I/D) forman parte del proyecto “Computación de Alto Desempeño: Arquitecturas, Algoritmos, Métricas de rendimiento y Aplicaciones en HPC, Big Data, Robótica, Señales y Tiempo Real”. En particular del sub-proyecto “Mode-

los y métodos computacionales. Procesamiento de señales y reconocimiento de patrones”.

1. Introducción

Los modelos matemáticos nos permiten predecir fenómenos naturales y el comportamiento de estructuras, dispositivos, procesos y sistemas diseñados por el hombre. Los métodos computacionales asociados a modelos determinísticos y probabilísticos permiten estudiar el dominio del problema considerando distintas condiciones estructurales y funcionales. El análisis aritmético computacional es esencial no sólo para el diseño de hardware específico sino para comprender el alcance y limitaciones efectivas de un determinado modelo en su implementación. A medida que el objeto o fenómeno real se aleja de las hipótesis fundamentales es necesario evaluar su validez y su utilidad respecto a los objetivos iniciales [1][2]. Este abordaje permite implementaciones en *hardware* o *software* que ofrezcan soluciones viables. Determinar si una solución es viable con fundamento científico, es una tarea compleja y dependiente del problema particular analizado [3]. Este proyecto tiene como primer objetivo analizar y proponer modelos computacionales, métodos y las soluciones particulares derivadas de los mismos [4]. La evaluación de rendimiento es un aspecto fundamental para poder validar las soluciones propuestas o los modelos analizados [5]. Por lo tanto, otro aspecto fundamental es el estudio de las métricas y pa-

radigmas de desempeño en sistemas específicos.

En la sección 2 se presenta un breve resumen de los temas de I/D específicos en el período actual. La sección 3 enumera resultados obtenidos y esperados. Finalmente, la sección 4 resume los objetivos con respecto a la formación de recursos humanos.

2. Líneas de Investigación

2.1 Modelos Probabilísticos

El primer objetivo de esta línea de investigación es estudiar si un determinado modelo es viable en el contexto de un sistema particular. El segundo objetivo general es determinar si es robusto cuando compromisos de diseño implican apartarse de las hipótesis iniciales del modelo.

2.1.1 Imágenes de Tiempo de Vuelo (ToF)

Las cámaras de ToF permiten obtener imágenes de rango, también denominadas $2 \frac{1}{2} D$. El ruido y los artefactos en este tipo de imagen requieren de modelos probabilísticos adecuados para caracterizar, filtrar y eventualmente reducir efectos indeseados. Los filtros no lineales útiles en otro tipo de imágenes requieren determinar sus parámetros críticos de forma particular en imágenes ToF. Segmentar imágenes de tiempo de vuelo requiere modelos de segmentación con características específicas [6][7][8]. En particular métodos estadísticamente robustos.

2.1.2 Datos Fuertemente Desbalanceados

Cuando el conjunto de patrones de entrenamiento de un clasificador probabilístico es marcadamente asimétrico, esta línea de investigación estudia el dominio específico del problema y la capacidad de generalización del sistema de clasificación. Se analizan paradigmas de aprendizaje automático estadístico supervisado y semi-supervisado [9][10][11].

2.2 Desempeño de Sistemas de Posicionamiento, Navegación y Vigilancia.

En los sistemas de posicionamiento, de navegación y de localización [12][13], el concepto de desempeño excede al habitual que está limitado a la calidad nominal de la estimación de ubicación y eventualmente a la confiabilidad [14][15]. En estos sistemas deben considerarse además los parámetros de integridad y continuidad que le garanticen al usuario que la información proporcionada por el sistema es correcta para que una operación crítica pueda realizarse en forma segura [16][17].

Un tema relacionado con el desempeño de los sistemas vigilancia es el volumen de transacciones con características aleatorias [18].

Otro tema de interés es la integración de sensores, no a través de la fusión de las estimaciones sino de la cooperación en etapas previas del procesamiento o en funciones particulares o en los límites del volumen de cobertura [19]. Esta línea de trabajo se avoca al estudio de problemas puntuales de desempeño en los sistemas mencionados, utilizando criterios y métodos diversos de modelado y procesamiento de señales [20].

2.3 Generación de Descriptores

El objetivo de esta línea de trabajo es mejorar la calidad de los descriptores obtenidos a partir de modelos probabilísticos y modelos espectrales, considerando: unicidad, invariancia, sensibilidad y su impacto en el sistema de clasificación.

2.4 Aritmética de Precisión Finita

El análisis de los errores debidos a la aritmética de precisión finita utilizada es necesario para asegurar sistemas robustos y confiables. El análisis de los errores de propagación y de la inestabilidad de un determinado sistema, forman parte de los objetos de análisis necesario para la implementación en *hardware* o *software* de los modelos propuestos. Por la importancia y por su carácter transversal respecto a las líneas de investigación del proyecto, es el tema

adoptado para evaluar, objetivamente, modelos y métodos particulares de enseñanza-aprendizaje.

3. Resultados y Objetivos

3.1 Resultados publicados

- Se estudiaron y propusieron métodos para detección en series temporales de fMRI [21][22].
- Se desarrollaron métodos de segmentación de imágenes de rango y supresión del plano de fondo [23][24][25][26].
- Se analizaron y propusieron alternativas para el agrupamiento de objetos de interés en video [27].
- Se estudió el desempeño de un método de exclusión de satélites en un sistema de ayuda a la aeronavegación basado en GNSS [28].
- Se presentaron resultados experimentales de un método de aprendizaje en aritmética computacional [29].
- Se propuso un método de segmentación espectral para imágenes ToF [30].
- Se estudió el comportamiento bajo carga de un algoritmo para programar transacciones de radares aeroportuarios [31].

3.2 Objetivos generales

- Desarrollar modelos y optimizar algoritmos particulares de clasificación supervisada y no supervisada.
- Evaluar métodos de análisis de desempeño y su aplicación sobre los clasificadores y conjuntos de datos particulares.
- Diseñar y evaluar técnicas híbridas de vigilancia SSR/ADS-B.
- Evaluar la monitorización de la integridad de los sistemas de ayuda a la navegación aérea basados en sistemas GNSS.

- Estudiar métodos de selección y extracción de características.
- Promover la interacción con otros grupos y líneas de I/D resultando en un mecanismo de permanente consulta y transferencia.

4. Formación de Recursos Humanos

La formación de recursos humanos en primer lugar implica la transferencia de los resultados de cada línea de investigación a las asignaturas de grado y cursos de postgrado que los integrantes dictan. Los alumnos también tienen la posibilidad de realizar trabajos supervisados de investigación, resultantes en tesis y tesis en el área. Debido al carácter trans-disciplinar de las líneas de I/D expuestas, se espera orientar y brindar apoyo a investigadores y alumnos de otras líneas de investigación relacionadas.

BIBLIOGRAFIA

1. Zaluzniak V. Essentials of Scientific Computing – Numerical Methods for Science and Engineering. Woodhead Publishing, 2008.
2. Juergen G. Coupled Systems: Theory, Models, and Applications in Engineering. CRC, 2014.
3. Torokhti A., Howlett P. Computational Methods for Modelling of Nonlinear Systems. Elsevier, 2007.
4. Gustafsson B. Fundamentals of Scientific Computing Springer, 2011.
5. Aslak T., et al. Elements of Scientific Computing. Springer, 2010.
6. Von Luxburg U. A Tutorial on Spectral Clustering. Statistics and Computing, v.17(4), 2007.
7. Kim H.Y., Giacomantone J. O., Cho, Z. H. Robust Anisotropic Diffusion to Produce Enhanced Statistical Parametric Map, Computer Vision and Image Understanding, v.99, pp.435-452, 2005.
8. Han Y., Feng X., Baciú G. Variational and PCA based natural image segmentation. Pattern Recognition 46, pp. 1971-1984, 2013.
9. Cortes C, Vapnik V, Support vector networks. Machine Learning v.20, pp.273-297, 1995.

10. Vapnik, V. *The Nature of Statistical Learning Theory*. N. Y. Springer, 1995.
11. Aytug H. Feature selection for support vector machines using Generalized Benders Decomposition. *European Journal of Operational Research*, v. 244(1), pp. 210-218, 2015.
12. Partap Misra, Per Enge. *Global Positioning System: Signals, Measurements and Performance*, Ganga-Jamuna Press, 2010.
13. Hakan Koyuncu, Shuang Hua Yang. *A Survey of Indoor Positioning and Object Locating Systems Indoor Positioning System*, *International Journal of Computer Science*, 2010.
14. Petevelo Mark. *Quantifying the performance of Navigation Systems and Standards for assisted-GNSS*, Inside GNSS, 2008.
15. Morurikis A., Roumeliotis S. *Performance Analysis of Multirobot Cooperative Localization*, IEEE, 2005.
16. Murphy T., et. al., *Fault Modeling for GBAS Airworthiness Assessments*, Navigation, 2012.
17. Cosmen-Schortmann J., Azaola-Sáenz, Martínez-Olagüe M. A., Toledo-López M., *Integrity in Urban and Road Environments and its use in Liability Critical Applications*, IEEE, 2008.
18. Pengfei Duan, Maarten Uijy De Haa. *Flight Test Results of a Measurement-Based ADS-B System for Separation Assurance*, Navigation, 2013.
19. Paolo Mariano, Patrizio De Marco, Claudio Giacomini, *Data Integrity Augmentation by ADS-B SSR Hybrid Techniques*, *Integrated Communications, Navigation and Surveillance Conference*, ICNS, 2018.
20. Sam Pullen, Todd Walter, Per Enge. *SBAS and GBAS Integrity for Non-Aviation Users: Moving Away from "Specific Risk,"* *International Technical Meeting of The Institute of Navigation*, 2011.
21. Giacomantone J., Tarutina T. *Diffuse Outlier Detection Technique for Functional Magnetic Resonance Imaging*. *Computer Science and Technology Series*. XVI Argentine Congress of Computer Science Selected Papers. pp. 255-265, 2011.
22. Giacomantone J., De Giusti A. *Detección de áreas de interés bajo la hipótesis de relación espacial de voxels activados en fMRI*. XX Congreso Argentino de Ciencias de la Computación. San Justo. Argentina, 2014.
23. Lorenti L., Giacomantone J. *Segmentación espectral de imágenes utilizando cámaras de tiempo de vuelo*. XI Workshop Computación Gráfica, Imágenes y Visualización. pp. 430-439. Mar del Plata, Argentina, 2013.
24. Lorenti L., Giacomantone J. *Time of flight image segmentation through co-regularized spectral clustering*. *Computer Science & Technology Series*. XX Argentine Congress of Computer Science. Selected papers. La Plata, Editorial de la Universidad Nacional de La Plata, 2015.
25. Giacomantone J., et al. *Supresión del plano de fondo en imágenes de tiempo de vuelo*. VII Workshop Procesamiento de Señales y Sistemas de Tiempo Real, 2016.
26. Lorenti, L., Giacomantone, J., Bria, O. N., De Giusti, A. E. (2017). *Fusión de información de geometría e intensidad para segmentación de imágenes ToF*. XXIII CACIC. La Plata, Argentina, 2017.
27. Lorenti L., Giacomantone J., De Giusti A. *Agrupamiento de trayectorias vía clustering espectral incremental*. XXII CACIC, pp. 222-231, 2016.
28. Bria, O., Giacomantone, J., Lorenti, L., *Excluding Ionospherically Unsafe Satellite Geometries in GBAS CAT-I*. XXII CACIC. CCIS 790: Revised CACIC Selected Papers, 790, pp. 243-252, 2018.
29. Giacomantone, J., Bria, O., *Proactive Independent Learning Approach: A case study in computer arithmetic*. XXII CACIC. La Plata, Argentina, 2017.
30. Lorenti L., Giacomantone J., Bria O., *Unsupervised ToF Image Segmentation through Spectral Clustering and Region Merging*. *Journal of Computer Science & Technology*, v.18(2), pp. 97-104, 2018.
31. Bria O., Giacomantone J., Villagarcía H., *Compound Interleaving Scheduling Algorithm for SLM Transactions in Mode S Surveillance Radar*. *Communications in Computer and Information Science* 995, pp. 297-312. Springer, 2019.

Redes de Sensores Inalámbricas y Simulación en Sistemas de Tiempo Real

Fernando Romero¹, Diego Encinas¹, Armando De Giusti^{1,2}, Santiago Medina¹, Lucas Maccallini, Martín Pi Puig¹, Horacio Villagarcía^{1,3}, Juan Manuel Paniego¹, Fernando G. Tinetti^{1,3}

¹Instituto de Investigación en Informática LIDI (III-LIDI)³
Facultad de Informática – Universidad Nacional de La Plata – Centro Asociado CIC

²CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas

³Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

{fromero, dencinas, degiusti, smedina, mpipuig, hvw, jmpaniego, fernando}@lidi.info.unlp.edu.ar
lucas.maccallini@gmail.com

Resumen

Esta línea de investigación se dedica al desarrollo y análisis de Sistemas de Tiempo Real. Dentro de esta temática se trabajó sobre tres sublíneas: 1) Redes de sensores inalámbricas. 2) Modelado y Simulación. 3) SOTR y hardware de comunicaciones utilizados en nodos de Sistemas Distribuidos de Tiempo Real

Dentro de las redes de sensores inalámbricas se trabajan diferentes escenarios de despliegue, tanto urbanos como rurales. En cuanto a las simulaciones se realizan tanto de sistemas de hardware y cloud computing como a situaciones tales como evacuaciones en casos de emergencia y transmisión de enfermedades.

Contexto

Esta línea de Investigación forma parte del proyecto 11/F024 – Computación de Alto Desempeño: Arquitecturas, Algoritmos, Métricas de rendimiento y Aplicaciones en HPC, Big Data, Robótica, Señales y Tiempo Real SubProyecto CAD-3.

Procesamiento para problemas de Tiempo Real / Robótica del Instituto de Investigación en Informática LIDI acreditado por la UNLP.

Palabras Claves: Tiempo Real, Simulación, Comunicaciones, Redes de Sensores, Microcontroladores, Cloud Computing.

1. Introducción

Los Sistemas de Tiempo Real (STR) se definen como aquellos que implican restricciones de tiempo en los plazos en que, ante una determinada entrada, deben producir una salida. Otra característica es que interactúan con el mundo físico, o sea sus entradas provienen de información producida por sensores y detectores y sus salidas están conectadas a actuadores [3] [4] [7] [8] [17]. Estas restricciones temporales a los plazos dependen fundamentalmente del medio físico a controlar. Para respetar esto debe haber una sincronización entre mundo físico o real y el procesamiento dentro del sistema de cómputo, que deberá contar con un reloj de tiempo real sincronizado con algún

estándar de tiempo físico. Por otro lado, los sensores pueden estar en una diferente locación física, a distancia, como también ser numerosos. Esto lleva a la necesidad de estudiar redes de sensores, tanto de conexión con conductores como inalámbricas. A veces estos sensores están montados sobre robots móviles, terrestres y aéreos [6] [14] [15] [23], en combinación con los sensores y actuadores. En el desarrollo de redes de sensores se utilizan placas de desarrollo basadas en microcontroladores como, Arduino, NodeMCU, CIAA [12] [11] [18] y Computadoras de Placa Simple como, Raspberry Pi, Raspberry Pi Zero W, utilizando diferentes SOTR (Linux RT-Preempt, FreeRTOS, MQX, OSEK-OS, Zephyr, Raspbian, etc.) [5]. Se realizan pruebas de alcance, integridad y funcionalidad de redes de sensores inalámbricas [19] [20] [21] [13] [33] principalmente utilizando módulos WiFi y LoRa [16]. También se estudiaron sistemas para conectar los nodos de una red a plataformas y servicios del Cloud [1] [13]

En el campo del modelado y simulación [10] [9] [15] [17], se las realiza para obtener datos que permitan predecir el comportamiento y la eficiencia de distintos sistemas ante diferentes escenarios. Además, se ajustan estas simulaciones con datos reales, lo cual permite luego realizar ensayos sobre la simulación: ejemplos de ellos son las simulaciones de robots, placas, incendios, evacuación de edificios en catástrofes y procesos industriales.

2. Resultados y Objetivos

Se han desarrollado tareas sobre los temas antes expuestos tales como:

- Construcción y estudio de redes de sensores inalámbricas basadas en WiFi y LoRa[1].

- Interconexión de diferentes Single Board Computers (SBC) a través de servicios (como brokers MQTT) para realizar tareas que requieren sincronización [24] [25] [26].
- Se evaluaron, desplegaron y compararon distintas plataformas de software dedicadas al Internet de las Cosas (IoT) [27] [28] [29] [30].
- Se experimentó con la interconexión de diferentes placas de desarrollo a través de un cloud público [24] [31] [32].
- Se realizó una comparación entre los sistemas operativos de tiempo real FreeRTOS y Zephyr [33][34] sobre placas con sistemas de comunicación wifi [18] [33]
- Modelado y simulación de arquitecturas de Cloud Computing para comparar con arquitecturas de HPC [21] [2].
- Modelado y simulación de transmisión de enfermedades intrahospitalarias [22].
- Modelado y simulación para el análisis de vida ictícola [23].
- Paralelización y aceleración en la ejecución de simulaciones.

3. Formación de Recursos Humanos

Se desarrollan trabajos de alumnos en la Convocatoria a Proyectos de Desarrollo e Innovación de la Facultad de Informática de la UNLP.

Además, se encuentran en desarrollo 2 Prácticas Profesionales Supervisadas (PPS) con las que concluyen sus estudios los alumnos de Ingeniería en Computación y Analista en TICs, orientadas a las redes de sensores. Y se concluyó una PPS con foco en la utilización de la plataforma IBM Watson.

De postgrado, investigadores del grupo están desarrollando un trabajo final de especialización, una tesis de Maestría y una tesis de Doctorado.

5. Referencias

- [1] Integration of Sensor Networks with Cloud ComputingS. Medina, F. Romero, and F. G. Tinetti, Short papers of the 8th Conference on Cloud Computing, Big Data & Emerging Topics (JCC-BD&ET 2020), ISBN: 978-950-34-1927-4, págs. 2-5, 2020.
- [2] R. Calheiros, R. Ranjan, A. Beloglazov, C. De Rose and R. Buyya "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms" Published online 24 August 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/spe.995.
- [3]Burns, A, A, Wellings. "Real-Time Systems and Programming Languages: Ada, Real-Time Java and C/Real-Time POSIX", Addison-Wesley Educational Publishers Inc., 2009.
- [4] Buttazzo, G. C., "Hard RealTime Computing Systems", Third edition, Springer, 2011.
- [5]"FreeRTOS - market leading RTOS (real time operating system) for embedded systems supporting 34 microcontroller architectures". [http:// www.freertos.org/](http://www.freertos.org/).
- [6]Jenkins, T., I. Bogost. "Designing for the internet of things: prototyping material interactions." In CHI'14 Extended Abstracts on Human Factors in Computing Systems, pp. 731-740. ACM, 2014.
- [7]Kopetz. H., "Real-Time Systems, Design Principles for Distributed Embedded Applications". Second Edition. Springer. 2011.
- [8]Liu, J. W. S. Liu, "Real Time Systems", Integre Technical Publishing Co., Inc., 2000
- [9]C. Macal, M. North, Tutorial on agent-based modeling and simulation part 2: how to model with agents, in: Proceedings of the Winter Simulation Conference, 2006.
- [10]PHILLIP A. LAPLANTE, SEPPO J. OVASKA. REAL-TIME SYSTEMS DESIGN AND ANALYSIS Tools for the PractitionerFourth Edition. A JOHN WILEY & SONS, INC., PUBLICATION. IEEE PRESS. 2012.
- [11]<http://www.proyecto-ciaa.com.ar/devwiki/doku.php?id=desarrollo:edu-ciaa:edu-ciaa-nxp>
- [12] <http://www.can-cia.de/can-knowledge/can/can-fd/>
- [13] Análisis de una plataforma de simulación para Cloud Computing. Un caso de estudio. Tomás Rosales, Julián Spinelli, Marcos Di Nardo, Román Bond, Daniel Rosatto, Diego Encinas, Fernando Romero XXVI Congreso Argentino de Ciencias de la Computación (CACIC 2020) La Matanza
- [14] F. G. Tinetti and O. C. Valderrama Riveros, "Unmanned Vehicles: Towards Heterogeneous Hardware Approaches," 2018 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2018, pp. 919-924.
- [15] Fernando G. Tinetti, Oscar C. Valderrama Riveros, Fernando L. Romero, "Unmanned Vehicles: Real Time Problems in Drone Receivers", Conf. on Computational Science & Computational Intelligence (CSCI'19), Las Vegas, Nevada, USA , 2019, pp. 1081-1085.
- [15] D. Black, SystemC: From the Ground Up. Second Edition, Springer, 2010.
- [16] LoRa <https://www.lora-alliance.org/> 2017
- [17] Proteus. <https://www.labcenter.com>. 2017
- [18] NodeMcu <http://www.nodemcu.com/> 2017

- [19] Akyildiz, Ian F., and Mehmet Can Vuran. "Wireless sensor networks" Vol. 4. John Wiley & Sons, 2010.
- [20] Lewis, Franck L. "Wireless sensor networks." *Smart environments: technologies, protocols, and applications* 11 (2004): 46.
- [21] Raghavendra, Cauligi S., Krishna M. Sivalingam, and Taieb Znati, eds. "Wireless sensor networks" Springer, 2006.
- [22] Maccallini, L. ., Encinas, D. O., & Romero, F. . (2021). An Approach to the Modeling and Simulation of Intra-Hospital Diseases. *Journal of Computer Science and Technology*, 21(2), e14. <https://doi.org/10.24215/16666038.21.e14>
- [23] Alcalá-Carrillo, M., Castillo-Vargasmachuca, S. G, & Ponce-Palafox, J. T. (2016). Efectos de la temperatura y salinidad sobre el crecimiento y supervivencia de juveniles de pargo *Lutjanus guttatus*. *Latin american journal of aquatic research*, 44(1), 159-164. <https://dx.doi.org/10.3856/vol44-issue1-fulltext-17>
- [24] <https://www.raspberrypi.com/documentation/computers/os.html>
- [25] <https://nodered.org/docs/>
- [26] <https://www.hivemq.com/>
- [27] <https://ubidots.com/>
- [28] <https://developers.mydevices.com/cayenne/features/>
- [29] <https://www.kaaiot.com/>
- [30] <https://thingsboard.io/>
- [31] <https://www.ibm.com/es-es/cloud/internet-of-things>
- [32] <http://www.raspbian.org/>
- [33] ESP32-DevKitC V4 Getting Started Guide. Disponible en internet en: <https://docs.espressif.com/projects/espressif/en/latest/esp32/hw-reference/esp32/get-start-ed-devkitc.html>
- [34] ESP32 Zephyr setup — Amarula Solutions's Wiki documentation. Disponible en internet: <https://wiki.amarulasolutions.com/zephyr/esp32/esp32-setup.html>. 20 de octubre de 2021.

ORCID autores:

Fernando Romero: 0000-0002-1498-3752

Diego Encinas: 0000-0002-6948-9786

A. De Giusti: 0000-0002-6459-3592

Santiago Medina: 0000-0001-6852-7165

Martín Pi Puig: 0000-0002-7202-7638

Horacio Villagarcía:

Juan Manuel Paniego: 0000-0001-6721-9822

Fernando G. Tinetti:

ESTUDIO DEL HABLA DE PACIENTES CON ENFERMEDAD DE PARKINSON Y DESARROLLO DE APLICACIÓN WEB

Monica Giuliano¹, Silvia N. Pérez¹⁻², Nahuel Mangiarua¹

¹Instituto de Ingenierías y Nuevas Tecnologías, Universidad Nacional del Oeste

²Departamento de Ingeniería e Investigaciones Tecnológicas –

Universidad Nacional de La Matanza

{mgiuliano; sperez}@uno.edu.ar

RESUMEN

En este trabajo se propone explorar y optimizar técnicas de procesamiento y análisis automático del habla para detectar y clasificar pacientes con enfermedad de Parkinson (EP). Además de desarrollar una aplicación web para interactuar con médicos y pacientes.

El proyecto tiene una línea de investigación que contempla dos líneas de trabajo. En la primera se considerarán audios de vocales, palabras, frases y habla espontáneas disponibles en bases de datos pública con habla de personas con y sin Parkinson. Se emplearán criterios estadísticos para optimizar los algoritmos para la detección y clasificación del grado de avance de la EP.

Paralelamente, se propone desarrollar una aplicación web que permita que los médicos puedan realizar grabaciones del habla de sus pacientes con EP en sus consultorios. Estas grabaciones formarán parte de una nueva base de datos para futuros análisis. Además, los audios requerirán una devolución de las características más importantes para médicos y pacientes, que se harán según el avance de la primera línea de trabajo.

Se espera como resultado ofrecer una interfaz para la interacción entre los investigadores, los desarrolladores informáticos y los médicos, que aporte al tratamiento de pacientes con EP.

Palabras clave: Parkinson, Procesamiento automático del habla, Aplicación web.

CONTEXTO

La propuesta que se presenta es parte del proyecto de investigación “Estudio del habla de pacientes con Enfermedad de Parkinson para la asistencia al diagnóstico y seguimiento” del Instituto de Ingenierías y Nuevas Tecnologías, Universidad Nacional del Oeste (UNO)

1. INTRODUCCIÓN

La enfermedad de Parkinson (EP) es un trastorno neurodegenerativo crónico y progresivo del sistema nervioso central, de comienzo gradual y progresión lenta, apareciendo en edades medias, entre 40 y 70 años. Las manifestaciones motoras de la EP están representadas por temblor de reposo, rigidez y bradicinesia [1]. En su evolución natural se agregan otros síntomas como: trastornos de la marcha, desequilibrio y alteraciones de la voz, que implican un marcado impacto en la calidad vida [2].

Se estima que entre un 60-80% de pacientes con EP presentan alteraciones de la voz, caracterizadas por cambios en frecuencia, duración e intensidad del habla [3].

Estas alteraciones se confunden en muchas ocasiones con los cambios naturales de los adultos mayores, en relación con la presbifonía (distensión y flacidez de los pliegues vocales) [4]. Otro síntoma frecuente en la EP es la presencia progresiva de disartria hipocinética (rigidez muscular e incapacidad de producir movimiento). La EP puede dejar marcas en la

distonía y el temblor de las cuerdas vocales. Los parámetros biomecánicos que controlan la tensión y el desequilibrio de las cuerdas vocales ayudan a rastrear la actividad neuromotora de las vías laríngea y articularia [5].

El análisis acústico de la voz permite detectar los cambios de los parámetros vocales, lo cual puede ser utilizado para predecir el avance de la EP y definir una intervención clínica específica [6]. Se ha observado que el análisis acústico podría actuar como marcador diagnóstico objetivo y no invasivo en la EP.

A modo de ejemplo se presentan en la Fig.1 ejemplos de la señal de fonación de la vocal /a/ para el caso de una persona sin EP en la primera imagen y con EP en la segunda.

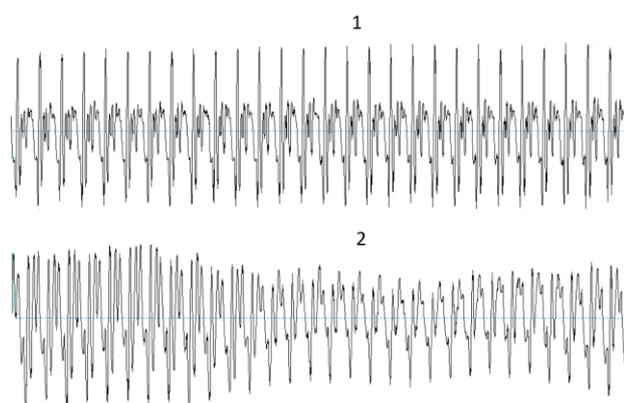


Fig. 1: 0.15 seg de la fonación de la vocal /a/, en (1) voz de persona sin y (2) con Parkinson.

Además, el diagnóstico y seguimiento de EP a través del análisis del habla pueden utilizar sistemas de tele-diagnóstico y tele-monitoreo, de bajo costo y de posible auto administración por parte de los pacientes [5, 7-9].

Los avances de los últimos años son variados y dispersos en cuanto a los mejores parámetros a estudiar con mayor desarrollo en el campo de la ingeniería. A nuestro criterio falta una mayor interacción con los médicos tratantes de EP y por ello creemos fundamental el aporte de la aplicación web (App) para profundizar este campo.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Se inicia con este proyecto una nueva línea de investigación para la UNO que incluye dos líneas de trabajo. La primera correspondiente al procesamiento de señales de audio con métodos automáticos. En segundo lugar, se trabaja en el desarrollo de una App que permita la interacción entre investigadores y médicos de pacientes con EP.

La App permitirá diferentes acciones desde la perspectiva de diferentes perfiles de usuarios: administrador, investigadores, médicos y pacientes. Los sistemas de software cliente-servidor con interfaces gráficas basadas en HTML, comúnmente llamadas App, conforman una alternativa realmente flexible a la problemática del despliegue y de la distribución de software. Mediante la estandarización de sus interfaces y protocolos de comunicación a aquellos de la web (HTTP, HTML, JSON, XML) las App puede ser utilizadas desde casi cualquier dispositivo actual con acceso a internet, facilitando el alcance y la distribución de la misma a lo largo y ancho de la geografía, atravesando múltiples estratos sociales. A su vez las tecnologías de “backend” basadas en java o .net otorgan una amplia capacidad de procesamiento y disponibilidad de bibliotecas de software de análisis.

En principio para este proyecto apuntamos a desarrollar una App en java utilizando spring/cdi sobre un servidor tomcat y JSF como frontend. No se descarta utilizar un framework de cliente para el frontend (angular) en vez de JSF si resulta conveniente para funcionalidades offline. A su vez, el backend java necesitará hacer interfaz con las herramientas y modelos de análisis que la otra línea de trabajo determine.

En el marco del proyecto se desarrollará una App que permita gestionar grabaciones de

habla de pacientes con EP para ser utilizada por los médicos en el consultorio

Becarios alumnos participarán del desarrollo de esta App. La app facilitará a un usuario personal médico la grabación y el almacenamiento de muestras de audio de pacientes que serán procesadas para generar y devolver distintos tipos de información que asistan al profesional de la salud.

Los administradores son informáticos formados y en formación, que trabajarán según los requerimientos de los demás usuarios.

Los investigadores buscan optimizar el análisis de la señal de voz a partir de la parametrización adecuada según los avances en el área, para luego aplicar los conocimientos a la App. El análisis de la voz de pacientes con EP requiere conocimientos en varias áreas: sistema fonatorio, procesamiento de señales, análisis estadístico y aprendizaje automático.

Se han utilizado varios métodos para la parametrización del habla de personas con EP. Se han considerado diversos algoritmos y metodologías para la selección de parámetros. La mayoría de los trabajos se dirigen a la fonación de la vocal /a/. A modo de ejemplo solamente se citarán algunos trabajos.

En el metanálisis realizado en [6], como revisión del análisis acústico de la voz en la enfermedad de Parkinson según distintos autores, se observó que no siempre se encuentran diferencias entre la condición de EP y no EP para las medidas de *Jitter* y *Shimmer*. Esto alienta a profundizar en el análisis de otros parámetros o sus variantes.

En [10] Tsanas se ha observado que combinando medidas de disfonías clásicas (*Shimmer*) con técnicas no lineales como EMD-ER (Empirical mode decomposition excitation ratio) o VFER (Vocal fold excitation ratio) y MFCC (Mel frequency cepstral

coefficients) de bajo orden, se obtiene un subconjunto de características que conduce a una más precisa evaluación de la fonación.

Las medidas de disfonía o parámetros que cuantifican las principales características que se observan en el habla de los parkinsonianos, se pueden considerar en cuatro grupos para su análisis [11-12]. El grupo 1 (tipo *Jitter*) y el grupo 2 (tipo *Shimmer*), que detectan problemas en la periodicidad, cuantificando variaciones de la frecuencia (F_0) y amplitud de la señal, respectivamente. El grupo 3 (tipo ruido) involucra problemas relacionados con el cierre incompleto de las cuerdas vocales, lo cual puede producir ruido. Finalmente, los parámetros del grupo 4 (MFCC) consideran problemas en la articulación de la fonación en el tracto vocal.

El aprendizaje automático o *machine learning* (ML) es un área de inteligencia artificial que utiliza el análisis por computadora para estudiar conjuntos de datos con el objetivo de detectar tendencias relevantes. En otras palabras, el aprendizaje automático enseña a las computadoras a aprender de los datos. Una rama importante de ML lo constituye el aprendizaje profundo o *deep learning* [13-14].

El análisis estadístico, sumado al ML, permite el trabajo criterioso con los datos de modo para comprender mejor los resultados y la caracterización del habla de los pacientes con EP.

El análisis automático del habla de personas con EP es una línea de investigación aún abierta que requiere jerarquizar las potencialidades del análisis del habla de pacientes con Parkinson utilizando modelos de inteligencia artificial para diagnóstico y seguimiento de la EP, sumando criterios físicos y estadísticos. A su vez los hallazgos debieran analizarse con especialistas de la salud, como médicos y fonoaudiólogos.

3. RESULTADOS ESPERADOS

El avance en el análisis del habla de personas con EP permitiría en el corto plazo brindar herramientas útiles, tanto a médicos y pacientes, que permitan el diagnóstico de alteraciones y seguimiento de la evolución de los trastornos en el habla.

Inicialmente se trabajará con bases de datos públicas, especialmente con una construida en Argentina [15]. Mas tarde, a partir de la App desarrollada se espera contar con una base de datos propia y creciente en volumen.

En la actualidad la telemedicina aumenta su proyección y efectividad. La utilización del análisis del habla a través de una App, ofrece a los médicos un elemento más de diagnóstico y seguimiento de la EP. Se han desarrollado otros software o aplicación [16-17], cuya utilización no ha prosperado entre los médicos argentinos de hospitales públicos

Se espera generar un espacio de construcción de conocimiento científico sobre el habla de paciente con EP, en base a la interacción entre investigadores de la UNO y personal de la salud de hospitales públicos argentinos, en primer lugar.

El análisis automático del habla de personas con EP requiere necesariamente de la conjunción interdisciplinaria de saberes, además de la identificación de innovaciones necesarias y posibles para el diagnóstico, seguimiento y tratamiento de los pacientes con EP en relación al habla.

Se busca desarrollar una herramienta informática que les permita a los médicos grabar las voces de los pacientes con EP y que, por ejemplo, les permita identificar tempranamente la necesidad de derivación a un tratamiento con fonoaudiólogos. La solución tecnológica estará representada por un desarrollo web que les pueda ser útil y accesible a los médicos

Se espera contribuir en la construcción de espacios de colaboración e intercambio multidisciplinar entre especialistas del campo de la salud (médicos, fonoaudiólogos, OTR, etc.) y tecnólogos (ingenieros, informáticos, matemáticos, etc.) que favorezca la generación de tecnologías que beneficien a personas con EP. Este espacio se iniciará en primer lugar con investigadores de la UNO y del Hospital Nacional Alejandro Posadas.

4. FORMACIÓN DE RECURSOS HUMANOS

La línea de investigación presentada colabora en la tesis doctoral de la Mg. Mónica Giuliano Además se promueve la formación de docentes de matemática y de becarios alumnos de la Escuela de Informática de la UNO.

Además, a través del proyecto se busca brindar servicios a hospitales públicos y médicos tratantes de la EP.

5. BIBLIOGRAFÍA

- [1] Jankovic, J. (2008). Enfermedad de Parkinson: características clínicas y diagnóstico. Revista de neurología, neurocirugía y psiquiatría, 79 (4), 368-376.
- [2] Kollensperger, M; Wenning, M.S.A.S.G. European, Red flags for multiples y stematopathy, Movement disorders: official journal of the Movement Disorder Society 23(8) (2008) 1093-9.
- [3] De Letter M, Santens P, Borsel, JV. (2003). The effects of levodopa on word intelligibility in Parkinson's disease. J CommunDisord 2005; 28: pp. 187-196.
- [4] Murray Morrison, Linda Rammage "Tratamiento de los trastornos de la voz" Elsevier España, 1996 capítulo 8, pág. 162.
- [5] Gómez-Vilda, P., Vicente-Torcal, M., Ferrández-Vicente, J., Álvarez Marquina, A., Rodellar-Biarge, V., Nieto-Lluis, V., et al. (2015a) Parkinson's disease monitoring from phonation biomechanics.

- [6] Chiaramonte, R., and Marco Bonfiglio. "Análisis acústico de la voz en la enfermedad de Parkinson: revisión sistemática de la discapacidad vocal y metaanálisis de estudios." *Revista de neurología* 70.11 (2020): 393-405.
- [7] Little, M.A, McSharry, P. E., Hunter, J. E. Spielman, J. & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, 56(4):1010–1022.
- [8] Orozco-Arroyave, J. R., Vásquez-Correa, J. C., Klumpp, P., Pérez-Toro, P. A., Escobar-Grisales, D., Roth, N., ... & Nöth, E. (2020). Apkinson: the smartphone application for telemonitoring Parkinson's patients through speech, gait and hands movement. *Neurodegenerative Disease Management*, 10(3), 137-157.
- [9] Tsanas, Athanasios, and Siddharth Arora. (2021). "Assessing Parkinson's Disease Speech Signal Generalization of Clustering Results Across Three Countries: Findings in the Parkinson's Voice Initiative Study." *Age* 63.10.8 (2021): 63.5.
- [10] Tsanas, M. A. Little, C. Fox and L. O. Ramig, (2014) "Objective Automatic Assessment of Rehabilitative Speech Treatment in Parkinson's Disease," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 1, pp. 181-190, Jan. 2014
- [11] Giuliano, M.; Fernandez L.; Pérez, S. (2020). Selección de Medidas de Disfonía para la Identificación de Enfermos de Parkinson. *Proceedings IEEE Congreso Bienal de Argentina (ARGENCON)*, 2020, pp. 1-8
- [12] Pérez, S. Fernandez L.; Giuliano, M.; (2021). Aporte de medidas de shimmer para la detección de enfermedad de Parkinson. *Jornadas de Avances en Métodos Estadísticos*, Colombia 2021.
- [13] Marsland. Stephen (2015). *Machine Learning: An Algorithmic Perspective – 2nd Ed.* CRC Press. 2015
- [14] Mitchell. Tom M. (1997). *Machine Learning.* WCB McGraw-Hill, 1997
- [15] Giuliano, M. Adamec. D, Debas, M. I. (2021). Construcción de una base de voz de personas con y sin enfermedad de Parkinson. *Revista Digital del Departamento de Ingeniería e Investigaciones Tecnológicas REDDI*. Vol.:6-Nro.1 (Julio-2021)
- [16] Orozco-Arroyave, J. R., Vásquez-Correa, J. C., Klumpp, P., Pérez-Toro, P. A., Escobar-Grisales, D., Roth, N., ... & Nöth, E. (2020). Apkinson: the smartphone application for telemonitoring Parkinson's patients through speech, gait and hands movement. *Neurodegenerative Disease Management*, 10(3), 137-157
- [17] Giuliano M Luciano E Massiolo R Procopio G Vaccotti F Viqueira. (2020a). Utilización de Software Parkinsoft para comparar voces de personas con y sin Enfermedad de Parkinson. *CONAISI 2020*. 05 y 06 de noviembre 2020 RIISIC – CONFEDI – UTN San Francisco.

Optimización de cálculo para la medición de la intensidad acústica compleja para dispositivos stand-alone de bajo costo

Carlos Esteban GRAFFIGNA¹ (cgraffigna@undec.edu.ar); Doménico STANZIAL² (domenico.stanzial@cnr.it); José Pablo DOÑA¹ (ingenierojpdg@gmail.com); Emanuel Frati¹ (fefrati@undec.edu.ar); Sebastián Guidet¹ (sguidet@undec.edu.ar)

¹Universidad Nacional de Chilecito
9 de julio 22, Chilecito, La Rioja, Argentina

²CNR Istituto di Microelettronica e Microsistemi
Bologna, Italia.

RESUMEN

En recientes publicaciones presentadas por la Universidad Nacional de Chilecito (UNdeC) en colaboración con el Consiglio Nazionale delle Ricerche (CNR) y la Università degli Studi di Ferrara (UNIFE) [Stanzial y Graffigna, 2016, 2017], donde se propone una sólida definición de la intensidad acústica compleja basada en la impedancia acústica, las tres instituciones han sido copropietarias de una patente sobre la metodología para la medición de la intensidad compleja del sonido en 3D en el dominio frecuencial.

Dado que la metodología propuesta implica un cálculo elevado y que se espera ejecutarla en un dispositivo stand-alone de bajo costo, es necesario optimizar los algoritmos de medición para maximizar el uso de los recursos de hardware disponibles.

Palabras claves: Cómputo paralelo, Dispositivo de medición, Intensidad sonora.

CONTEXTO

El proyecto nace a partir de una patente desarrollada durante una beca doctoral cofinanciada entre el Consiglio Nazionale delle Ricerche (CNR) de Italia, la Università degli Studi di Ferrara (UNIFE) de Italia y la Universidad Nacional de Chilecito (UNdeC) de Argentina.

Una vez concluido satisfactoriamente el doctorado, se firma un nuevo acuerdo de

colaboración científica entre la UNdeC y el Istituto per la Microelettronica e Microsistemi (IMM) perteneciente al CNR, el cual tiene como objetivo principal promover la investigación científica y tecnológica en el campo de la acústica energética, de sus microsistemas sensores y de sus aplicaciones prácticas ambientales e industriales. Además, prevé la cooperación en actividades de formación de personal con tesis de grado y posgrado, intercambio de docentes e investigadores, elaboración de publicaciones científicas y documentos técnicos, entre otras.

La invención consiste en una metodología para medir la intensidad compleja del sonido en 3D en el dominio frecuencial. Dado que es necesario ejecutar los algoritmos en un dispositivo stand-alone para tomar mediciones en campo, se implementaron los algoritmos en un dispositivo Raspberry Pi 3B+ obteniendo resultados importantes, pero con tiempos de procesamiento lentos.

Dado que la UNdeC está trabajando en un proyecto de tecnología HPC [1] se incluyó el mismo como una línea de investigación para optimizar los algoritmos de medición utilizando procesamiento paralelo para mejorar los tiempos de cálculo en la Transformada Rápida de Fourier y las operaciones matriciales.

INTRODUCCIÓN

La energética acústica estudia el comportamiento de la energía sonora dentro

de un campo acústico, el cuál es definido a partir de dos variables: presión sonora (escalar) y velocidad de vibración de la partícula acústica (vectorial) [2].

Si bien esta rama de la acústica viene siendo estudiada desde hace ya medio siglo [3,4], en los últimos años ha tenido un gran impulso debido al desarrollo de nuevos sensores y técnicas para la medición de la velocidad de la partícula acústica.

En la actualidad, existen sensores termo-resistivos que permiten la medición directa de la velocidad de la partícula sonora (sonda p-v) y con la evolución de los micrófonos de tecnologías MEMS ha sido posible realizar otras sondas de bajo costo que miden dicha velocidad de manera indirecta a través del gradiente de presión (sonda p-p).

La medición de estas dos variables contemporáneamente permite conocer entre otras cosas el campo de energía total, cinética y potencial, la velocidad de propagación de la energía, la cantidad de energía absorbida y reflejada dentro del sistema acústico, y la intensidad sonora [5].

Respecto a esta última, existe una definición clara sobre la intensidad activa (parte de la intensidad relacionada al flujo efectivo de la energía sonora que viaja en el medio), en cambio, la definición de intensidad reactiva todavía es un problema controversial.

En recientes publicaciones presentadas por la UNdeC en colaboración con el CNR (Italia) [7,8,9], se ha propuesto una sólida definición de la intensidad acústica compleja (activa y reactiva), basada en la impedancia acústica, a partir de la cual la UNdeC es co-propietaria de una patente que describe la metodología para la medición del espectro de la intensidad compleja del sonido en las 3 dimensiones espaciales [9], el tensor de reactividad espectral, el vector de intensidad activa frequency-overall, el módulo de la intensidad aparente frequency-overall, y la reactividad total del sistema.

Si bien los algoritmos de medición han sido satisfactoriamente ejecutados y verificados en ambiente de laboratorio utilizando equipos performantes y software de cálculo numérico especializado, es necesario desarrollar un sistema stand-alone capaz de procesar las señales acústicas en campo.

Como primera aproximación, se elaboró un prototipo para la medición de las variables energéticas utilizando un dispositivo Raspberry Pi3B+ [6]. Los algoritmos fueron traducidos en Python y se desarrolló una interfaz gráfica elemental para operar.

Dado que los recursos de hardware disponible en estos dispositivos son significativamente menores a los utilizados con laboratorio, los tiempos de procesamiento, tanto en el análisis frecuencial como en el cálculo de matrices, son altos en relación a la dinámica de los sistemas acústicos.

Por este motivo, el presente proyecto persigue como objetivo principal optimizar los algoritmos para mejorar los tiempos en la medición de la intensidad compleja del sonido y demás variables energéticas, maximizando el aprovechamiento del hardware disponible a través del uso de la GPU.

LÍNEAS DE I+D

Para medir la intensidad acústica compleja [7,8] es necesario aplicar transformaciones de Fourier para obtener los espectros frecuenciales, y calcular otras variables físicas de tipo escalar, vectorial y tensorial (matriz) en el dominio de los números complejos [9].

En base a los resultados obtenidos del primer prototipo, se determinó que los procesos con mayor carga computacional se dan en la Transformación de Fourier y en el cálculo matricial para la obtención del tensor de reactividad en el dominio de la frecuencia.

De este análisis se plantean las siguientes líneas de investigación para optimizar el proceso de medición:

Análisis FFT en GPU

Si bien existen distintos trabajos relacionados a la programación utilizando la GPU de los dispositivos de Raspberry y otros donde se utiliza programación paralela para la transformación de Fourier, es difícil encontrar publicaciones que utilicen la GPU de los Raspberry para aplicar la transformación de Fourier utilizando librerías estándar de Python.

Esta línea de investigación propone desarrollar los algoritmos de procesamiento paralelo para procesar la Transformada de Fourier de señales temporales, utilizando la GPU disponible en los dispositivos Raspberry, e integrarla en librerías estándar de Python como scipy.

Calculo matricial en GPU

Dado que uno de los cálculos que emplea mayor tiempo en la medición de la intensidad compleja de sonido es el cálculo del tensor de reactividad en el dominio de la frecuencia, se pretende desarrollar un algoritmo paralelo que permita optimizar los tiempos de cálculo utilizando la GPU de los dispositivos Raspberry.

Si bien el cálculo matricial ejecutado por el algoritmo es relativamente simple, el volumen de procesamiento viene dado por la cantidad de matrices que deben ser calculadas. Esto se debe a que se debe computar un tensor de reactividad por cada componente frecuencial.

FORMACIÓN DE RECURSOS HUMANOS

Este proyecto surge como resultado de una tesis doctoral en Física con especialización en Acústica, donde el doctorando realizó el primer prototipo funcional utilizando librerías estándares de python [6].

Actualmente está en curso una Tesis de Maestría que pretende cubrir la primera línea de investigación que busca optimizar el cálculo de la Transformada de Fourier paralelizando el algoritmo y ejecutándolo en la GPU de la Raspberry Pi.

BIBLIOGRAFÍA

- [1] F. E. Frati y otros, "Tecnología HPC como motor de ciencia de la UNdeC", XXIII Workshop de Investigadores en Ciencias de la Computación, Chilecito, 2021.
- [2] Dudley H Towne. *Wave phenomena*. Addison-Wesley, Massachusetts (USA), 1967.
- [3] Frank Fahy. *Sound intensity*. CRC Press, 1989.
- [4] Finn Jacobsen. *Springer Handbook of Acoustics, chapter 25. Sound Intensity, pages 1053–1075*. Springer-Verlag New York, 2014.
- [5] D. Stanzial, Giorgio Sacchi, and Giuliano Schiffrer. "On the physical meaning of the power factor in acoustics". *J. Acoust. Soc. Am.*, 2012.
- [6] C. Graffigna, *Measuring Complex Sound Intensity from Wave Impedance*. Tesis doctoral. Universidad Nacional de Chilecito, La Rioja, Argentina. Universidad de Ferrara. Ferrara Italia, 2019.
- [7] D. Stanzial y C. E. Graffigna. *On the connection between wave impedance, sound intensity and kinetic energy in monochromatic fields*. *Proceedings of Meetings on Acoustics*. Buenos Aires, 2016.
- [8] D. Stanzial y C. E. Graffigna. *On the general connection between wave impedance and complex sound intensity*. *Proceedings of Meetings on Acoustics*. Boston, 2017.
- [9] D. Stanzial y C. E. Graffigna. *Precision device for measuring the three-dimensional spectra of complex intensity*. *Proceedings of Meetings on Acoustics*. Boston, 2017.

Técnicas de big data para el análisis de la teoría moderna de optimización de carteras de inversión en el marco del mercado argentino de valores

Rodrigo Gonzalez^{1,2}, Santiago Eguren¹, y Carlos A. Catania^{1,3}

¹ Universidad Champagnat, Facultad de Informática y Diseño,
Mendoza, Argentina.
{gonzalezrodrigo,cataniacarlos}@uch.edu.ar

² Universidad Tecnológica Nacional, Facultad Regional Mendoza, GridTICs,
Mendoza, Argentina.

³ Universidad Nacional de Cuyo, Facultad de Ingeniería, LABSIN, Mendoza,
Argentina.
harpo@ingenieria.uncuyo.edu.ar

RESUMEN

La teoría moderna de configuración de carteras de inversión (Modern Portfolio Theory, MPT) propone diseñar una cartera que proporcione la máxima rentabilidad asumiendo una cantidad de riesgo determinada. En general, la MPT ha sido fuertemente analizada en mercados de países industrializados. El mercado argentino de valores (MAV) es un mercado cuya volatilidad está en sintonía con la de los mercados internacionales, pero que también responde a circunstancias propias. Además, a menudo la inflación no es tenida en cuenta cuando se analiza el desempeño de una cartera de inversión en mercados bursátiles de países industrializados. Contrariamente, esta variable debe tenerse en cuenta cuando se desea formar una cartera de inversión con activos del MAV. La ciencia de datos es un campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de un sistema a partir de sus datos. Además, provee herramientas para la visualización conveniente de todos los aspectos relevantes sobre el sistema que se trata de modelar. Por tanto, este proyecto propone analizar el comportamiento del MAV en los últimos 10 años con técnicas del análisis cuantitativo (quantitative analysis) y de ciencia de los datos, para determinar

diferentes carteras de inversión óptimas con activos argentinos para distintos niveles de riesgo, según la teoría moderna de configuración de carteras de inversión.

Palabras Claves: Análisis cuantitativo, Ciencia de datos, Modern Portfolio Theory, Cartera de inversión, Lenguaje R, Python.

CONTEXTO

El presente proyecto se desarrolla en el marco de la Facultad de Informática y Diseño. Este trabajo es parte de las líneas de investigación llevadas adelante por el Dr. Rodrigo Gonzalez junto al Dr. Carlos A. Catania desde el año 2021.

1. INTRODUCCIÓN

La teoría moderna de configuración de carteras de inversión (Modern Portfolio Theory, MPT) argumenta que es posible diseñar una cartera óptima que proporcione a un inversor la máxima rentabilidad asumiendo una cantidad de riesgo determinada. La MPT fue desarrollada por el economista Harry Markowitz (Markowitz, 1952). Esta teoría destaca la importancia de las carteras de inversión, el riesgo, la diversificación y las conexiones entre los diferentes tipos de activos. Los activos enfrentan tanto riesgos sistemáticos (riesgos de mercado) como

riesgos no sistemáticos (características específicas de cada activo). La diversificación adecuada de una cartera no puede prevenir el riesgo sistemático, pero puede amortiguar, si no eliminar, el riesgo no sistemático. En general, la MPT y sus teorías derivadas (Markowitz, 1959) (Rockafellar, 2000) (Low, 2016) han sido fuertemente analizadas en mercados de países industrializados (Konno, 1991) (Schanzenbach, 2017).

El mercado argentino de valores (MAV) es un mercado cuya volatilidad está en sintonía con la de los mercados internacionales, pero que también responde a circunstancias propias, como la gran caída que sufrió el índice Merval en agosto de 2019. Por otra parte, en mercados de países industrializados generalmente los bonos de un país son considerados instrumentos de baja volatilidad. En el caso del MAV, los bonos del estado argentino han mostrado en los últimos años una volatilidad alta cuando se los compara con instrumentos similares de otros países (Rodríguez, 2017). Además, en el contexto del MAV es necesario tener presente los efectos de la inflación en una cartera de inversión (Spotorno, 2008) (Dapena, 2013) (Brito, 2018). En general, la inflación no es tenida en cuenta cuando se analiza el desempeño de una cartera de inversión en mercados bursátiles de países industrializados. Por lo expuesto, el MAV presenta ciertas características que lo hacen particular.

El MAV puede considerarse formado por dos grandes mercados financieros, ByMA (ByMA, 2020) y ROFEX (ROFEX, 2020), donde el primero se especializa en acciones y bonos, y el segundo en contratos de futuros y opciones. Así, el MAV está compuesto por acciones nacionales y del mercado de los EE.UU. (CEDEAR), bonos nacionales y provinciales, bonos corporativos (ON, obligaciones negociables), contratos de futuros y opciones (puts y calls) sobre acciones, bonos, commodities (cereales, carne vacuna, petróleo, metales), cotización del dólar (dólar futuro) e índices (ROFEX20). Si se considera la cotización de todos estos

activos en los últimos 10 años, se estima que el total de instrumentos de inversión puede llegar a 500. Por tanto, para hacer un análisis exhaustivo para la confección de una cartera de activos óptima en el MAV tomando los cotizaciones de los últimos 10 años, es necesario utilizar herramientas que permitan acceder, analizar, almacenar y visualizar una gran cantidad de información. Respecto a los datos requeridos, estos no están concentrados en una sola fuente, por lo que es necesario recurrir a diferentes APIs (Application Programming Interface) para acceder a datos de los mercados ByMA y ROFEX. Además, es posible que algunos datos necesarios para evaluar una cartera de activos no estén disponibles a través de una API, como la serie de tiempo de algún con la cotización de activo en particular o balances de empresas cotizantes en el MAV, por lo que será necesario implementar técnicas de scraping para extraer datos extras en formato PDF, Excel o HTML de forma automática.

La ciencia de datos es un campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, ya sea estructurados o no estructurados. Además emplea técnicas y teorías extraídas de muchos campos dentro del contexto de las matemáticas, la estadística, la ciencia de la información, y la informática. Es una continuación de algunos campos de análisis de datos como la estadística, minería de datos y aprendizaje automático. Se la define como "un concepto para unificar estadísticas, análisis de datos, aprendizaje automático, y sus métodos relacionados, a efectos de comprender y analizar los fenómenos reales". La ciencia de datos es una herramienta muy utilizada en finanzas. Permite manejar la gran cantidad de datos que proveen los mercados financieros diariamente, y analizar las intrincadas relaciones que pueden existir entre diferentes activos. Además, provee herramientas para la visualización conveniente de todos los datos relevantes

sobre el fenómeno financiero que se desea modelar.

Por lo expuesto, este proyecto propone analizar el comportamiento del MAV en los últimos 10 años con técnicas del análisis cuantitativo (quantitative analysis) y la ciencia de los datos, para determinar diferentes carteras de inversión óptimas con activos argentinos para distintos niveles de riesgo, según la teoría moderna de configuración de carteras de inversión.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Para el desarrollo del presente proyecto pueden diferenciarse 3 etapas principales.

1. Entender los alcances y limitaciones de la teoría moderna de carteras de inversión. Se procederá a estudiar las bases teóricas de la MPT. Se usará la bibliografía más relevante en el área y se tomarán cursos online sobre manejo de carteras de inversión en sitios destacados como Coursera y EdX. Además, algunos integrantes del equipo de trabajo propuesto estudiarán los lenguajes de programación R y Python, para estar al mismo nivel que el resto de los investigadores.

2. Crear un entorno computacional para adquisición, análisis, almacenamiento y visualización de datos de mercados de valores. Se creará un entorno computacional compuesto por funciones y algoritmos derivados del estudio y análisis de la MPT. Se utilizarán activos del mercado de valores estadounidense, ya que la gran mayoría de la bibliografía existente sobre MPT utiliza este mercado para su estudio. Además, hay abundante documentación sobre distintas fuentes de información (bibliotecas en Python y R) que permiten acceder a estos datos. Los datos de mercado se guardarán en una base de datos centralizada. La misma estará compuesta por los precios de acciones, bonos, contratos de futuros y opciones, e información financiera adicional asociada a estos instrumentos como índices dividendos de empresas, alfa, beta, TIR, duration, exit yield, entre otros. Como consecuencia de la diversidad de las fuentes de datos y su

probable falta de estructura, se prevé que se deberán aplicar diferentes técnicas de scraping para sitios online, archivos PDF y tablas de Excel para extraer información sobre estados contables de empresas cotizantes o sus posibles reparto de dividendos. Se utilizarán herramientas de visualización para grandes volúmenes de datos.

3. Determinar diferentes carteras de inversión óptimas según diferentes niveles de riesgo para el mercado argentino de valores entre los años 2010 y 2020. En esta última etapa se determinarán las carteras de inversión óptimas, según la MPT, para el mercado argentino de valores al tomar datos de los últimos 10 años. Se definirá la frontera eficiente (efficient frontier) para diferentes combinaciones de activos, nivel de riesgo y ganancias esperadas. Todos estos valores deberán ser ajustados por inflación y transformados a dólares estadounidenses, para poder comparar directamente con las carteras de inversión encontradas en el mercado estadounidense. Se deberán estudiar los mecanismos de accesos al MAV que ofrecen tanto ByMA como ROFEX.

3. RESULTADOS ESPERADOS

Al término del tiempo de duración del plan de trabajo, se espera obtener los siguientes resultados:

1. Sólida comprensión de los beneficios y limitaciones de la teoría moderna de creación de carteras de inversión.
2. Creación de un entorno computacional para el análisis y la visualización de carteras de inversión con información de los mercados de valores argentino y estadounidense.
3. Mejor entendimiento de las particularidades del mercado argentino de valores, respecto al mercado de valores de los EE.UU.
4. Obtención de diferentes tipos de carteras de inversión para el mercado argentino, según el nivel de riesgo que desee tomar un inversor en particular. Además, los resultados obtenidos serán publicados tanto en conferencias internacionales de alta calidad con referato, como en revistas especializadas de alto

impacto, parte del catálogo de Scopus y Web of Science.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo de este proyecto está integrado por:

1. El Dr. Rodrigo Gonzalez, codirector de este proyecto, se especializa en procesamiento digital de señales, análisis de series de tiempo, modelado de sistemas, estadística y procesos estocásticos. Cuenta con experiencia en técnicas de aprendizaje automático para el modelado de sistemas utilizando series de tiempo (Gonzalez and Catania, 2019).

2. El Dr. Catania, docente e investigador de la Universidad Nacional de Cuyo y de la Universidad Champagnat, propuesto como investigador formado, es un especialista en el área de Ciencia de los Datos (Data Science). Cuenta con experiencia en la resolución de diferentes problemas aplicando algoritmos específicos del área de aprendizaje automático (Catania et al., 2018) (Gonzalez and Catania, 2019).

3. El Mg. Eguren es Licenciado en Economía y Magíster en Administración de Negocios (MBA), ambos títulos otorgados por la Universidad Nacional de Cuyo. Además, es docente de la cátedra de "Estadística Aplicada a la Logística" en la Facultad de Ciencias Económicas de la Universidad Nacional de Cuyo. Es un especialista en análisis cuantitativo y teoría bursátil. Además, cuenta con experiencia práctica en programación en los lenguajes de programación Python y R.

En el campo de la formación de recursos humanos, se espera:

1. Iniciar en investigación a profesionales que estén interesados en postularse a diferentes becas doctorales, en el marco del proyecto en referencia.

2. Iniciar la investigación a alumnos, tanto de la UCH, UTN o UNCuyo, que deseen sumarse al proyecto como becarios.

3. Iniciar en investigación a alumnos de los últimos años de carrera que deseen desarrollar

su proyecto final de carrera en el marco de este proyecto.

5. BIBLIOGRAFÍA

(Brito, 2018) Mateo Brito y Ronen Goldrossen. Instrumentos de cobertura contra la inflación argentina. Tesis de grado. Departamento de Economía. Universidad de San Andrés. Diciembre de 2018.

(ByMA, 2020) Bolsa y Mercados Argentinos S.A. CABA, Argentina. Sitio web: <https://www.byma.com.ar/>.

(Dapena, 2013) José P. Dapena. La evolución del índice Merval (precio de acciones) y la inflación. Newsletter CEA, Dpto. Finanzas. Universidad del Cema. Marzo de 2013.

(Konno, 1991) Konno, H. and H. Yamazaki (1991). "Mean-Absolute Deviation Portfolio Optimization Model and Its Application to Tokyo Stock Market." *Management Science*. Vol. 37, No. 5, May 1991, pp. 519–531.

(Low, 2016) Low, R.K.Y.; Faff, R.; Aas, K. (2016). "Enhancing mean-variance portfolio selection by modeling distributional asymmetries". *Journal of Economics and Business*. 85: 49–72. doi:10.1016/j.jeconbus.2016.01.003.

(Markowitz, 1952) Markowitz, H. "Portfolio Selection." *Journal of Finance*. Vol. 7, No. 1, March 1952, pp. 77–91.

(Markowitz, 1959) Markowitz, H. M. *Portfolio Selection: Efficient Diversification of Investments*. John Wiley & Sons, Inc., 1959.

(Rockafellar, 2000) Rockafellar, R. T. and S. Uryasev. (2000) "Optimization of Conditional Value-at-Risk." *Journal of Risk*. Vol. 2, No. 3, Spring 2000, pp. 21–41.

(Rodríguez, 2017) Rodríguez, J. N., Bueno, J. G., Barrios, G. R., & Díaz, S. Diversificación internacional de portafolio en los mercados accionarios de Argentina, Brasil, Chile, Colombia, México y Perú. En el Congreso Internacional en Administración de Negocios Internacionales (CIANI 2017). Septiembre de 2017, págs. 355-375.

(ROFEX, 2020) ROFEX S.A. Ciudad de Rosario, Santa Fe, Argentina. Sitio web: <https://www.rofex.com.ar/>.

(Schanzenbach, 2017) Schanzenbach, M. M., & Sitkoff, R. H. The prudent investor rule and market risk: an empirical analysis. *Journal of Empirical Legal Studies*, 14(1), 129-168.

(Spotorno, 2008) Fausto Spotorno y Alejandro Javier Aranda. Portafolio de cobertura contra la inflación en Argentina. Tesis de Maestría. Universidad del Cema. Diciembre de 2008.

(Catania et al., 2018) Carlos A. Catania; Sebastian Garcia; Pablo Torres. An Analysis of Convolutional Neural Networks for detecting DGA. Argentina. La Plata. 2018. Libro. XXIV Congreso Argentino de Ciencias de la Computación 2018. Red de Universidades con Carreras en Informática.

(Gonzalez and Catania, 2019) Rodrigo Gonzalez and Carlos A. Catania. Time-delayed multiple linear regression for de-noising MEMS inertial sensors. *Computers & Electrical Engineering*, Volume 76, 2019, Pages 1-12, ISSN 0045-7906.

RCCI – Redes de Cooperación Científica Internacionales

Formación de recursos humanos orientados a Ciudades Inteligentes Sostenibles– Proyecto CAP4CITY

Armando De Giusti , Patricia Pesado , Ariel Pasini ,
Pablo Thomas , Rocío Muñoz 
Elsa Estevez , Pablo Fillotrani , Sonia Rueda, Karina Cenci, Gabriela A. Diaz

Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata
50 y 120 - La Plata, Buenos Aires
Centro Asociado CIC
526 e/ 10 y 11 - La Plata, Buenos Aires
(degiusti, ppesado, apasini, pthomas, rmunoz) @lidi.info.unlp.edu.ar

Laboratorio de Ingeniería de Software y Sistemas de Información (LISSI)
Departamento de Ciencias e Ingeniería de la Computación – Universidad Nacional del Sur
Av. San Andrés 800 – Campus de Palihue - Bahía Blanca, Buenos Aires
Centro Asociado CIC
526 e/ 10 y 11 - La Plata, Buenos Aires
(ece, prf, svr, kmc, gabriela.diaz) @cs.uns.edu.ar

RESUMEN

El proyecto CAP4CITY tiene como objetivo fortalecer y desarrollar la capacidad de instituciones académicas en América Latina y Europa para mejorar la calidad de la educación superior en el campo de ciudades inteligentes sostenibles. El Proyecto es co-financiado en el marco del Programa Erasmus+ de la Unión Europea, para la construcción de capacidades en la educación superior. Se presenta el avance de los primeros años del proyecto y las actividades a realizar en el siguiente periodo.

Palabras Claves

Ciudades Inteligentes - Capacidades de Gobernanza - Gobernanza Digital - Colaboración Internacional

CONTEXTO

La línea de investigación y desarrollo aquí descrita es parte de las tareas planificadas por el proyecto “Strengthening Governance Capacity for Smart Sustainable Cities (CAP4CITY)”. El Proyecto, financiado como parte del programa Erasmus+ de la Unión Europea

es ejecutado por un consorcio integrado por 12 universidades; cuatro de ellas europeas – Donau Universität für Weiterbildung (DUK) en Austria, Tallinn University of Technology (TUT) en Estonia, Delft University of Technology (TU Delft) en los Países Bajos, y Gdańsk University of Technology (GUT) en Polonia; y ocho universidades en la región de América Latina – Universidad Nacional de La Plata (UNLP) y Universidad Nacional del Sur (UNS) en Argentina; Pontificia Universidade Católica do Rio Grande do Sul (PUCRS) y Faculdade Meridional (IMED) en Brazil; Universidad Técnica Federico Santa María (UTFSM) y Universidad Católica del Norte (UCN) en Chile; y Universidad Externado de Colombia (UEC) y Escuela Colombiana de Ingeniería (ECI) en Colombia. El Proyecto lleva el número 598273 y se ejecuta bajo el acuerdo 598273-EPP-1-2018-1-AT-EPPK A2-CBHE-JP. Debido a la pandemia de COVID-19 el proyecto ha sido extendido, teniendo una duración de 48 meses.

1. INTRODUCCION

El proyecto CAP4CITY tiene como principal objetivo utilizar la gran atención

que el concepto de Ciudades Inteligentes Sostenibles (SSC, por sus siglas en inglés) ha alcanzado en América Latina e integrarlo en varios cursos universitarios utilizando nuevas herramientas de enseñanza y aprendizaje, así como desarrollar nuevos planes de estudio en todos los niveles del proceso educativo.

Se define una SSC como una ciudad innovadora que utiliza Tecnologías de Información y Comunicación (TIC) y otros medios para mejorar la calidad de vida, la eficiencia de la operación y los servicios urbanos, y la competitividad, al tiempo que se garantiza que satisfaga las necesidades de las generaciones presentes y futuras con respecto a los aspectos económicos, sociales, ambientales y culturales.

Uno de los modelos de SSC propone cinco dimensiones para su desarrollo: 1) Social, 2) Económica, 3) Ambiental, 4) Gobernanza, y 5) Infraestructura Urbana. La Figura 1 muestra el modelo.

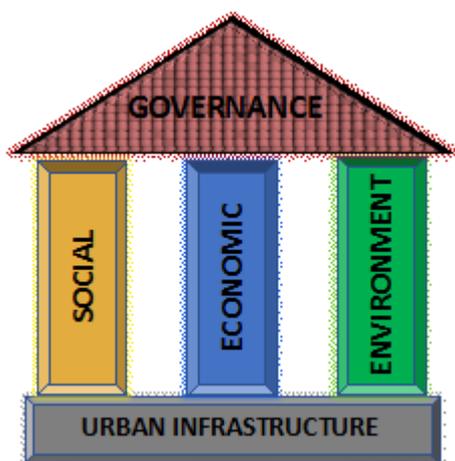


Figura 1. Dimensiones de una CIS

La dimensión Social cubre los aspectos relacionados con las personas y comunidades para garantizar la calidad de vida, como, por ejemplo: salud, seguridad, educación, entre otros. La Económica cubre los aspectos relacionados con el crecimiento económico responsable y sustentable, y la generación de oportunidades laborales. Por su parte, la Ambiental se relaciona con la utilización

de prácticas ecológicas, la protección y restauración del medio ambiente. La dimensión de Gobernanza se refiere a la capacidad de administrar recursos, políticas e involucrar a diferentes partes interesadas, proponiendo mecanismos y procesos regulatorios y de cumplimiento bien equilibrados de manera estandarizada y continua. Por último, la dimensión de Infraestructura Urbana se refiere a la infraestructura física (carreteras, transporte, etc.) y a la infraestructura digital (tecnología de la información y comunicación), que son herramientas esenciales para permitir ciudades inteligentes y sostenibles.

Dadas estas cinco dimensiones, la complejidad de las mismas, así como la característica multidisciplinaria de los problemas a resolver en SSC, hace que la construcción de capacidades humanas, particularmente, la formación de profesionales que se dediquen al liderazgo y gobernanza de estas iniciativas sea un gran desafío.

Debido al creciente número de competencias necesarias y su característica interdisciplinaria, los planes de estudio de SSC se implementarán en áreas como Administración de Empresas, Informática, Ingeniería, Arquitectura y Urbanismo, Planificación Urbana, Ciencias Políticas, entre otras, a través de una red de colaboración internacional de instituciones académicas en América Latina y Europa, que apoyan la modernización e internacionalización del campo de la educación superior.

A fin de definir las competencias necesarias para el desarrollo de SSC en América Latina, se hizo un relevamiento de datos en la región. A tal efecto, se realizaron dos talleres en las ciudades de Bogotá y Medellín, Colombia, como experiencia piloto. Luego cada una de las universidades de Latino América replicó el *“Taller de Relevamiento de Competencias*

para el Desarrollo de Ciudades Inteligentes y Sostenibles” con representante de su comunidad, incluyendo ciudadanos y representantes de la industria, la academia y el gobierno.

Con los resultados obtenidos de todas las universidades se estableció una lista de competencias en base a las cuales se definieron 31 cursos. Estos cursos fueron validados mediante entrevistas online, encuestas y workshops con los interesados locales y se contó, además, con el asesoramiento de expertos internacionales externos al proyecto, obteniendo así una retroalimentación de los cursos en general, de cada una de las áreas establecidas y de los contenidos que serán impartidos.

Los beneficiarios del Proyecto CAP4CITY incluyen instituciones de educación superior a nivel local, provincial, nacional e internacional. Estas podrán utilizar los cursos diseñados, compartir conocimientos e intercambiar experiencias sobre el desarrollo y la entrega de programas educativos relacionados con SSC.

El resto de este artículo se organiza de la siguiente manera: la sección 2 explica los objetivos de investigación, la sección 3 los resultados esperados/obtenidos y, por último, la sección 4 discute la formación de recursos humanos.

2. OBJETIVOS DE INVESTIGACIÓN Y ACTIVIDADES

- Mejorar la calidad de la educación superior en el campo de SSC, aumentando su relevancia para el mercado laboral y la sociedad en general.
- Aumentar las competencias de los recursos humanos en los países en desarrollo de América Latina para enfrentar los desafíos del mundo digital.

- Proveer el desarrollo continuo de competencias en SSC a través de programas de capacitación y educación.
- Asegurar una enseñanza multidisciplinaria orientada a la resolución de problemas en SSC.
- Desarrollar una red de cooperación entre los socios de diferentes regiones del mundo.
- Promover la colaboración entre los socios del Consorcio, las entidades públicas, las empresas y otros interesados en SSC.
- Facilitar el intercambio de conocimientos, experiencias y buenas prácticas e iniciativas conjuntas en materia de SSC entre socios académicos ubicados en diferentes ciudades y países.

La Universidad Nacional de La Plata y la Universidad Nacional del Sur colaborarán con las siguientes actividades:

- Relevamiento de iniciativas de SSC que puedan identificarse como buenas prácticas y de programas de postgrado relacionados con temas de SSC.
- Capacitación de formadores en SSC.
- Implementación de nuevos programas educativos en SSC.
- Difusión y explotación del concepto de SSC en Argentina.
- Fortalecimiento de los vínculos laborales con sectores públicos y privados de la región.

3. RESULTADOS ESPERADOS/OBTENIDOS

Al primer trimestre de 2022, los resultados obtenidos por la UNLP y la UNS son:

- Se ha desarrollado un relevamiento de programas de postgrado relacionados con temas de SSC.

- Se organizaron talleres para identificar competencias y validar los módulos de formación propuestos.
- Se han diseñado y validado 5 de los 31 cursos (de diferentes niveles – pregrado, posgrado y formación continua) relacionados a SSC.
- Se han estudiado y desarrollado dos casos de estudio relacionados con SSC en Argentina.
- Se ha desarrollado e implementado un curso a distancia de “capacitación de formadores” sobre la construcción de MOOC para los cursos del proyecto.
- Se grabaron distintos MOOC para los cursos desarrollados y para los casos de estudio en Argentina.
- Se dictaron cursos de posgrado relacionados con los contenidos desarrollados sobre Ciudades Inteligentes Sostenibles en la UNLP.
- Se trabajó en la definición de una nueva Maestría Profesional conjunta en Ciudades Inteligentes Sostenibles que será ofrecida y dictada en colaboración entre la UNLP y la UNS.
- Se está organizando un evento de difusión a nivel nacional para presentar los resultados del proyecto.
- Desarrollo de tesis de postgrado y tesinas de grado en el área.
- Participación de los integrantes de esta línea de investigación en el dictado de asignaturas/cursos de grado/postgrado en la Facultad de Informática de la UNLP y en el Departamento de Ciencias e Ingeniería de la Computación de la UNS.
- Fortalecimiento de la colaboración institucional internacional entre las universidades miembros del consorcio.

5. AGRADECIMIENTOS

Project co-funded by the Erasmus+ Programme of the European Union. Grant no: 598273-EPP-1-2018-1-AT-EPPKA2-CBHE-JP.

6. BIBLIOGRAFIA

- [1] Akande, A., Cabral, P., Gomes, P., Casteleyn, S. 2019. “The Lisbon ranking for smart sustainable cities in Europe.” *Sustainable Cities and Society*, 44, pp. 475-487.
- [2] Caragliu, Andrea, Chiara Del Bo, and Peter Nijkamp. 2011. “Smart Cities in Europe.” *Journal of Urban Technology* 18(2):65–82.
- [3] Craglia, Massimo, Lila Leontidou, Giampaolo Nuvolati, and Jürgen Schweikart. 2004. “Towards the Development of Quality of Life Indicators in the ‘Digital’ City.” *Environment and Planning B: Planning and Design* 31(1):51–64.
- [4] Elsa Estevez, Nuno Lopes, Tomasz Janowski. 2016. “Smart sustainable cities: Reconnaissance Study”, 1-330.
- [5] Huovila, A., Bosch, P., Airaksinen, M. 2019. “Comparative analysis of standardized indicators for Smart sustainable cities: What indicators and standards to use and when?” *Cities*, 89, pp. 141-153.
- [6] Keshvardoost, S., Renukappa, S., Suresh, S. 2019. “Developments of policies related to smart cities: A critical review”. *Proceedings -*

4. FORMACION DE RECURSOS HUMANOS

Los esfuerzos dedicados a la formación de recursos humanos incluyen:

- Capacitación de los miembros del proyecto en SSC.

- 11th IEEE/ACM International Conference on Utility and Cloud Computing Companion, UCC Companion 2018, art. no. 8605807, pp. 365-369.
- [7] Kurebayashi, Toshihiko, Yoshihiro Masuyama, Kiyonori Morita, Naoyuki Taniguchi, and Fumio Mizuki. 2011. "Global Initiatives for Smart Urban Development." *Hitachi Review* 60(2):89–93.
- [8] Martin, C., Evans, J., Karvonen, A., Paskaleva, K., Yang, D., Linjordet, T. 2019. "Smart-sustainability: A new urban fix?" *Sustainable Cities and Society*, 45, pp. 640-648.
- [9] Paroutis, Sotirios, Mark Bennett, and Loizos Heracleous. 2014. "A Strategic View on Smart City Technology: The Case of IBM Smarter Cities During a Recession." *Technological Forecasting and Social Change* 89:262–72.

SI – Seguridad Informática

Consortio para la colaboración en I+D+I en Temas de Cloud Computing, Big Data y Emerging Topics (CCC-BD&ET)

**Proyecto Integrador: “Transformación Digital en la incorporación de la Resiliencia como
un Key Performance Indicator de Prestaciones Sociales (KPIS)”**

III-LIDI – Instituto de Investigación en Informática LIDI (UNLP – Argentina)

LISSI–Laboratorio de Investigación y Desarrollo en Ingeniería de Software y Sistemas de Información (UNS – Argentina)

VyGLab – Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (UNS– Argentina)

LIDIC – Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (UNSL – Argentina)

HPC4EAS – High Performance Computing for Efficient Applications and Simulation (UAB – España)

SMILe – Soft Management of Internet and Learning (Universidad de Castilla-La Mancha – España)

ArTeCS – Group of Architecture and Technology of Computing Systems (UCM – España)

LITRP – Laboratorio de Investigaciones Tecnológicas en Reconocimiento de Patrones (UCM – Chile)

LCG – Laboratorio de Computación Gráfica (UNSL – Argentina)

ITIC – Instituto para las Tecnologías de la Información y las Comunicaciones (UNCu – Argentina)

DisCo – Grupo de I+D+I en Computación Distribuida (Universidad de Zaragoza – España)

& Investigadores Asociados al CCC- BD&ET

Resumen

El Consorcio de I+D+i en *Cloud Computing, Big Data & Emerging Topics* (CCC-BD&ET) [1] es una iniciativa para fomentar y formalizar la colaboración existente entre grupos de investigación de varias universidades en temáticas vinculadas a Cloud Computing, al Análisis Masivo de Datos y a Tópicos Emergentes, como las tecnologías 4.0, entre otros. Estas temáticas, y su integración, han adquirido creciente importancia por su aplicación en dominios de alto impacto como las ciudades inteligentes, la internet de las cosas, los sistemas de e-health y los basados en tecnologías de block-chain.

Los integrantes del consorcio, provenientes mayoritariamente de Argentina, Chile y España han tenido, a lo largo de los años, diversas experiencias de trabajo conjunto que fueron consolidadas a partir de la organización y realización de las *Jornadas de Cloud Computing-Big Data & Emerging Topics* (JCC-BD&ET) llevadas a cabo en la Universidad Nacional de La Plata (Argentina). La constitución de este Consorcio, reafirma y formaliza estas líneas de colaboración proponiendo acciones de cooperación

académica vinculadas con la formación de recursos humanos, la formulación y ejecución de proyectos conjuntos, y la vinculación con empresas y organismos relacionados con la industria informática, entre otras.

Este trabajo presenta el avance del consorcio en la definición de un proyecto integrador que tiene como eje la Resiliencia para la Transformación Digital.

Palabras clave: Prestaciones Sociales, Resiliencia, Objetivos de Desarrollo Sostenible, Cloud Computing, Big Data, HPC, Data Analytics, Sistemas Inteligentes, Emerging Technologies.

Contexto

El Consorcio de I+D+i en *Cloud Computing, Big Data & Emerging Topics* (CCC-BD&ET) resulta de la cooperación, llevada a cabo durante varios años, entre grupos de investigación, desarrollo e innovación de Universidades vinculadas con las realización anual de las *Jornadas de Cloud Computing-Big Data & Emerging Topics* (JCC-BD&ET), organizadas por la Universidad Nacional de La Plata (UNLP), en Argentina.

I. Introducción

Las JCC-BD&ET surgieron como *Jornadas de Cloud Computing* realizándose por primera vez del 17 al 19 de junio de 2013 en La Plata, organizadas por el III-LIDI, Facultad de Informática, de la UNLP. A partir del 2015, pasaron a ser las *Jornadas de Cloud Computing & Big Data* y, en el año 2020, se constituyeron en JCC-BD&ET.

Estas Jornadas se llevan a cabo anualmente en la Facultad de Informática de la UNLP y constituyen un foro de intercambio de ideas, proyectos, resultados científicos y aplicaciones concretas en diferentes áreas relacionadas con Cloud Computing, Inteligencia de Datos, Big Data y Tecnologías Emergentes. Desde sus inicios, en el contexto de las Jornadas se desarrollan conferencias, paneles, cursos de posgrado, y también se integran ponencias científicas con experiencias de desarrollos y aplicaciones, fomentando la interacción entre la academia y los sectores productivos/industriales. A partir de estos encuentros fueron surgiendo distintas actividades en colaboración, que no sólo se formalizaron mediante acuerdos entre los distintos actores. Los resultados de las colaboraciones se encuentran reflejados en las publicaciones detalladas en las páginas mencionadas en las referencias.

Como corolario de las distintas actividades de colaboración realizadas por los participantes a las JCC-BD&ET es que surge la propuesta de conformar un Consorcio de I+D+i en Cloud Computing, Big Data & Emerging Topics que permita afianzar y proyectar a futuro las relaciones existentes [1]. En este contexto de colaboración, las JCC-BD&ET seguirán constituyendo un foco anual de encuentro, independientemente del trabajo académico y científico que se realice durante el año para la concreción de los objetivos del Consorcio.

Objetivos

Uno de los objetivos del Consorcio es formular un proyecto de investigación que integre las capacidades de los grupos de investigación que lo componen y que a la vez sea un punto de partida para la formulación de posibles presentaciones conjuntas de financiamiento en áreas relacionadas [2].

El Consorcio propone avanzar con una visión prospectiva que dé respuestas a problemas o desafíos estructurales de la sociedad del futuro. De este modo, no sólo se limita a definir un objetivo sino también a delinear una metodología de trabajo que compromete a nuestras áreas de investigación y desarrollo a abordar con rigor científico y con mirada social el camino de la transformación digital que está atravesando nuestra sociedad del siglo XXI.

Esta metodología se enfocará en las posibilidades de incidencia de algunos de los temas transversales de dominio del Consorcio en diferentes retos de los Objetivos de Desarrollo Sostenible (ODS) [3] y en la aplicación de herramientas y estrategias de la Transformación Digital.

II. Descripción del proyecto integrador

Los ODS constituyen una iniciativa de la ONU y se definen como 17 retos cuyo fin es erradicar la pobreza, proteger el planeta y garantizar que todas las personas del mundo, sin distinción, gocen de paz y prosperidad.

La Transformación Digital es un proceso que surge como una alternativa capaz de producir un cambio o una innovación en una organización a partir de la incorporación inteligente de diferentes tecnologías digitales en todos los niveles y funciones de ésta.

Todo modelo de innovación propone como punto de partida el conocimiento y la asimilación de lo que ya se conoce del problema planteado. Luego, en base a las posibles soluciones, se incorpora un proceso de evaluación y comparación de alternativas, que debe hacerse a través de ciertos valores medibles que demuestren efectivamente cómo el sistema está alcanzando los objetivos seleccionados. Este proceso introduce una métrica denominada “*Key Performance Indicator* (KPI)” cuya especificación depende del ámbito de aplicación.

En el proyecto se trabajará sobre Índices/Indicadores específicos relacionados con prestaciones sociales (KPIS) y su impacto sobre los Objetivos de Desarrollo Sostenible (ODS).

A modo de ejemplo, la siguiente tabla muestra ámbitos transversales que forman parte de la *expertise* del Consorcio y los retos seleccionados de los ODS en los que puede impactar una transformación digital.

Aplicaciones de interés social	ODS
En Educación	4,5,11 y 16
Gobierno Digital	5,11,15 y 16
Industriales y Agrícolas	9 y 15
En Economía	11,15 y 16
En Salud	4,5 y 11
En Ciudades Inteligentes y Sostenibles	4,11 y 16
Móviles y su evolución a las Web progresivas (WPA)	4,9 y 16

En todos los casos, el Consorcio propone diversas tecnologías para lograr sus objetivos, dado que una transformación digital no subyace en una única herramienta sino que

existen múltiples procesos claves involucrados para afrontar la transformación:

Tecnologías de la transformación digital	Herramientas
Datos, Información y Conocimiento	Modelos y Simulación - IA/Inteligencia de Datos Big Data - Análisis Visual - Realidad Extendida - Resiliencia
Computación Avanzada	Cloud Computing – HPC, Arquitectura y Aceleradores - IoT – Edge/Fodge Computing

Finalmente, la metodología presentada en este artículo, propone incorporar la capacidad del sistema u organización bajo transformación para soportar y sobreponerse a circunstancias de adversidad y desastres, como un KPI de las prestaciones sociales (KPIS). Este indicador, denominado Resiliencia, evaluará la capacidad de recuperar un estado operativo cuando ha cesado la perturbación a la que había estado sometido.

III. Resultados esperados

Los resultados esperados están relacionados con el fortalecimiento de las capacidades de los grupos de investigación del Consorcio, en términos de sus RRHH, redes de trabajo, acceso a recursos organizacionales y mejoramiento de la calidad de los resultados producidos, producto de la colaboración interdisciplinaria e inter-universitaria promovida por el Consorcio. La definición de un proyecto integrador ordena y permite integrar los aportes de los participantes del proyecto.

IV. Formación de RRHH

En lo concerniente a la formación de recursos humanos los distintos integrantes están abocados a la formación de recursos humanos tanto a nivel de grado como de posgrado, en temáticas afines al Consorcio y al Proyecto en particular que en muchos casos se realiza en colaboración entre distintos grupos participantes.

V. Integrantes del Consorcio

El Consorcio está integrado tanto por grupos de investigación como por investigadores asociados. A continuación se detallan los mismos.

Grupos que conforman el CCC- BD&ET

III-LIDI – Instituto de Investigación e Informática LIDI (Universidad Nacional de La Plata – Argentina) [4]
Directora: *Lic. Patricia Pesado*

LISSI – Laboratorio de Investigación y Desarrollo en Ingeniería de Software y Sistemas de Información (Universidad Nacional del Sur – Argentina) [5]
Director: *Dr. Pablo Fillottrani*

VyGLab – Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (Universidad Nacional del Sur – Argentina) [6]
Directora: *Dra. Silvia Castro*

LIDIC – Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (Universidad Nacional de San Luis – Argentina) [7]
Director: *Dr. Marcelo Errecalde*

HPC4EAS – High Performance Computing for Efficient Applications and Simulation (Universidad Autónoma de Barcelona – España) [8]
Director: *Dr. Emilio Luque*

SMILe – Soft Management of Internet and Learning (Universidad de Castilla-La Mancha – España) [9]
Director: *Dr. José A. Olivas Varela*

ArTeCS – Group of Architecture and Technology of Computing Systems (Universidad Complutense de Madrid – España) [10]
Director: *Dr. Francisco Tirado*

LITRP – Laboratorio de Investigaciones Tecnológicas en Reconocimiento de Patrones (Universidad Católica de Maule – Chile) [11]
Representante para el Consorcio: *Dr. Ricardo Barrientos*

LCG – Laboratorio de Computación Gráfica (Universidad Nacional de San Luis – Argentina) [12]
Director: *Mg. Roberto Guerrero*

ITIC – Instituto para las Tecnologías de la Información y las Comunicaciones (Universidad Nacional de Cuyo – Argentina) [13].
Director: *Dr. Carlos García Garino*

DisCo – Grupo de I+D+I en Computación Distribuida (Universidad de Zaragoza – España) [14].
Coordinador: *Dr. Pedro Javier Álvarez Pérez-Aradros*

Investigadores Asociados al CCC- BD&ET

Dr. Aurelio Fernández (Universidad Rovira i Virgili – España).
Dr. Emmanuel Frati (Universidad Nacional de Chilecito – Argentina).
Dr. Javier Balladini (Universidad Nacional de Comahue – Argentina).
Dra. Adriana Gaudiani (Universidad Nacional de General Sarmiento – Argentina).
Dra. Mónica Denham (Universidad Nacional de Río Negro – Argentina).
Lic. Nelson Rodríguez (Universidad Nacional de San Juan – Argentina).

Referencias

- [1] Consorcio de I+D+I en *Cloud Computing, Big Data & Emerging Topics* (CCC-BD&ET), <https://jcc.info.unlp.edu.ar/consorcio-de-idi-en-cloud-computing-big-data-emerging-topics/>.
- [2] Consorcio para la colaboración en I+D+I en temas de Cloud Computing, Big Data y Emerging Topics (CCC-BD&ET)- WICC 2021. 978-987-24611-3-3; 978-987-24611-4-0. Páginas: 778-783.
- [3] La UNESCO y los Objetivos de Desarrollo Sostenible: <https://es.unesco.org/sdgs>
- [4]<http://weblidi.info.unlp.edu.ar>
- [5]<https://lissi.cs.uns.edu.ar>
- [6]<http://vyglab.cs.uns.edu.ar>
- [7]<http://lidic.unsl.edu.ar>
- [8] <https://grupsderecerca.uab.cat/hpc4eas/>
- [9]<http://smile.esi.uclm.es>
- [10]<https://artecs.dacya.ucm.es>
- [11]<http://www.litrp.cl>
- [12]<http://www.lcg.unsl.edu.ar/>
- [13] <http://itic.uncu.edu.ar/>
- [14] <http://webdiis.unizar.es/DISCO/>

Contratos Inteligentes para Internet de las Cosas

Jorge Eterovic; Marcelo Cipriano; Luis Torres; Dalma Agostina Lomoro

Instituto de Investigación en Ciencia y Tecnología
Dirección de Investigación Vicerrectorado de Investigación y Desarrollo.
Universidad del Salvador.
Lavalle 1854 – C1051AAB -Ciudad Autónoma de Buenos Aires - Argentina

{jorge.eterovic; cipriano1.618 }@gmail.com; torreslu@ar.ibm.com; agostina.lomoro@usal.edu.ar

RESUMEN

Internet de las Cosas (IoT) es un concepto que se refiere a interconectar distintos dispositivos a través de Internet, lo que puede traer muchos beneficios a la sociedad de diferentes maneras. Como los dispositivos están conectados en diferentes contextos y dominios, la información que se genera y se transmite involucra a múltiples partes interesadas.

Esta información puede variar desde lo privado y confidencial hasta lo público, por lo que es muy importante investigar y proponer soluciones para asegurar la integridad de origen, la seguridad y la interoperabilidad. Esto constituye un gran desafío.

En este trabajo de investigación se discute cómo los contratos inteligentes (Smart Contracts) y la tecnología de la cadena de bloques (Blockchain) pueden, potencialmente, llegar a ser una solución viable, aunque todavía no se ha encontrado una manera eficiente y segura para vincular todos los dispositivos de IoT con los contratos inteligentes, ya que no todos los objetos de IoT tienen la potencia computacional necesaria para implementarlo.

El resultado esperado es encontrar soluciones basadas en contratos inteligentes que mejoren la seguridad y la gestión de la información, identificando nuevas oportunidades y desafíos, y brindando recomendaciones y pautas de seguridad para los datos y las comunicaciones de los dispositivos interconectados.

Palabras Clave:

Contratos Inteligentes. Ethereum. Blockchain. Internet de las Cosas.

CONTEXTO

El Vicerrectorado de Investigación y Desarrollo (VRID), perteneciente a la Universidad del Salvador (USAL), dicta las políticas referidas a la investigación, concibiéndola como un servicio a la comunidad y entendiendo que los nuevos conocimientos son la base de los cambios sociales y productivos. Con el impulso de las propias Unidades Académicas se han venido desarrollando acciones conducentes a concretar proyectos de investigación uni/multidisciplinarios, asociándolos a la docencia de grado y postgrado y vinculando este accionar, para potenciarlo, con otras instituciones académicas del ámbito nacional e internacional.

La Dirección de Investigación, dependiente del VRID, brinda soporte a las distintas Unidades de Investigación y a sus investigadores para el desarrollo de Proyectos y Programas de Investigación, nacionales e internacionales, como así también, apoyo y orientación de recursos para la investigación.

A ella pertenece el Instituto de Investigación en Ciencia y Tecnología (RR 576/12) en el cual se enmarca este proyecto denominado “Integración de Blockchain e Internet de las Cosas usando Contratos Inteligentes.”, con una duración de 2 años (2021-2022) y que ya ha sido evaluado y aprobado para su realización.

1. INTRODUCCIÓN

Si buscamos una tecnología que impactará y beneficiará nuestras vidas en los próximos

años, es el Internet de las cosas. Los automóviles, electrodomésticos, teléfonos inteligentes, medidores de servicios públicos, sensores incorporados al cuerpo, indumentaria y casi cualquier cosa que podamos imaginar estarán conectados a Internet y serán accesibles desde cualquier parte del mundo [1]. La revolución que generará IoT será inigualable, algunos autores dicen que será similar a la construcción de carreteras y ferrocarriles que impulsaron la Revolución Industrial de los siglos XVIII al XIX [2], y será transversal a todos los sectores de la sociedad y todas las industrias, desde educación, salud, hogar y ciudad inteligente, hasta manufactura, minería, comercio, logística y vigilancia, solo por mencionar algunas [3].

En Internet de las cosas está involucrada directa o indirectamente la generación de cantidades significativas de información. Un grupo dinámico de partes interesadas debe tener distintos niveles de derechos de acceso a esa información. Además, el alcance de la información relacionada con IoT variará según los requisitos del dominio de la aplicación y el contexto de los dispositivos.

Las aplicaciones de IoT involucran a muchas partes interesadas, con diferentes roles y funcionalidades que acceden a distintos tipos de información con varios niveles de acceso, identidades múltiples y condiciones particulares de seguridad para cada una de ellas. Administrar todos estos activos de manera eficiente, segura e interoperable es un problema desafiante. Se analizará si la tecnología Blockchain y los contratos inteligentes pueden desempeñar un papel importante en este sentido [4].

Una cadena de bloques mantiene una colección, o libro mayor, de transacciones de manera descentralizada y distribuida. El libro mayor es inmutable e irreversible, lo que significa que las transacciones pasadas no pueden ser modificadas por ninguna entidad que registre transacciones en la Blockchain, y se comparte y sincroniza en todos los nodos participantes. De esta manera, la cadena de bloques garantiza que el libro mayor no puede

ser manipulado, y que todos los datos que posee la Blockchain son confiables [5].

Una cadena de bloques puede ser pública [6] o estar restringida solo a usuarios autorizados [7]. La Blockchain se considera una forma democrática de mantener transacciones [8] y se prevé que proporcione mecanismos de seguridad novedosos, que contribuyan a la sostenibilidad de las aplicaciones de IoT y permitan nuevos modelos de confianza [9].

Un contrato inteligente es una aplicación distribuida que vive en la cadena de bloques [6]. Esta aplicación es, en esencia, una clase de lenguaje de programación con campos y métodos. Los usuarios pueden interactuar con los campos y métodos públicos de esta clase enviando transacciones a su dirección en la cadena de bloques.

Cada vez que un usuario interactúa con un contrato inteligente, todos los nodos de la red Blockchain ejecutan todas las operaciones de manera determinista y confiable y uno de estos nodos se selecciona para almacenar el resultado de la ejecución de los contratos, si corresponde, en la cadena de bloques. Los contratos inteligentes pueden verificar las identidades y firmas digitales de los usuarios de la Blockchain, realizar cálculos de propósito general e invocar a otros contratos [10].

El código de un contrato inteligente es inmutable y no puede ser modificado ni siquiera por su propietario [11]. Además, todas las transacciones enviadas a un contrato se registran en la cadena de bloques, por lo que es posible obtener todos los valores históricos de una variable del contrato.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La línea de investigación propone analizar cómo se pueden usar los contratos inteligentes y la tecnología Blockchain para proporcionar mecanismos básicos de seguridad, facilitar la gestión de la información y permitir la interacción con los dispositivos de manera interoperable.

Los contratos inteligentes permiten que las aplicaciones interactúen con los dispositivos de IoT de manera similar a cómo los controladores de hardware permiten que las aplicaciones interactúen con los dispositivos de hardware. Es decir, los contratos inteligentes pueden describir las capacidades de un dispositivo, los servicios que ofrece y cómo se puede acceder a él.

Al escribir los contratos inteligentes, los desarrolladores deberían poder integrar los dispositivos en sus sistemas y procesos, para ofrecer servicios innovadores y sostenibles. Además, al aprovechar los anclajes de confianza proporcionados por los contratos inteligentes y la cadena de bloques subyacente, debería ser más fácil crear aplicaciones distribuidas abiertas, confiables y seguras, así como nuevos mecanismos de seguridad, responsabilidad y cobro. Del mismo modo, con los contratos inteligentes, los usuarios pueden realizar fácilmente transacciones con sus monedas digitales o incluso con sus tokens personalizados.

Una propiedad interesante de los contratos inteligentes es que son deterministas y siempre se ejecutan correctamente, por lo tanto, el propietario de un contrato no debe preocuparse si la lógica de la aplicación, por ejemplo, una condición de control de acceso, incluida en el contrato será respetada o no. Además, el código de un contrato generalmente está disponible y el propietario de un contrato puede incluso agregar controles adicionales en un contrato para proteger a sus clientes. Incluso de sí mismo. Por ejemplo, puede implementar un compromiso de dos fases para los pagos: un cliente, en lugar de pagar directamente a un proveedor de servicios, compromete algo de dinero o activos digitales a un contrato; estos fondos se mantienen en depósito por el contrato hasta que el proveedor de servicios proporcione el servicio esperado; en caso de que no lo haga, los fondos se devuelven al cliente.

Finalmente, los contratos inteligentes no se pueden eliminar de una cadena de bloques. Esta propiedad es útil para construir mecanismos de comunicación resistentes al

repudio (non-repudiation). Por ejemplo, y como veremos a continuación, un contrato inteligente puede proporcionar un puntero a la ubicación de un elemento de información, o incluso al elemento en sí. Si esta ubicación se incluye, de alguna manera, en una lista negra, el puntero puede actualizarse fácilmente por el titular del contrato. De manera similar, un contrato inteligente puede contener metadatos que pueden usarse para prevenir fraudes y elementos de contenido falso.

El código fuente del contrato, así como el valor de cada campo del contrato, está disponible públicamente. Por esta razón, los desarrolladores de contratos deben tener mucho cuidado al decidir qué información se almacenará en un contrato. En muchos casos es preferible almacenar los datos fuera de la cadena de bloques y almacenar en la cadena de bloques el hash de los datos. De esta manera, un usuario puede verificar la integridad de los datos recibidos. Además, y por la misma razón, se recomienda a los desarrolladores que utilicen herramientas estáticas (SAST) y dinámicas (DAST) que analicen automáticamente el código del contrato y detecten posibles fallas y riesgos de seguridad [12].

Cabe señalar que cuando decimos que los dispositivos interactúan con una cadena de bloques o con un contrato inteligente, siempre se da a entender que esto es a través de una puerta de enlace confiable. Además, el término usuario puede referirse a entidades del mundo real, aplicaciones, dispositivos o incluso contratos inteligentes.

Se analizan algunos modelos de interacción para IoT basados en contratos inteligentes:

Push-Pull: En este modelo de interacción, los dispositivos hacen que los datos, tanto el contenido como la información, estén disponibles y los empujen (Push) para que los usuarios pueden extraerlos. Para la operación de empuje se pueden considerar los siguientes enfoques:

- a) Los dispositivos empujan los datos en la cadena de bloques. Los datos se almacenan en la cadena de bloques, posiblemente a

través de un contrato inteligente. y se pueden usar contratos inteligentes para recuperarlos.

- b) Los dispositivos empujan los datos a los nodos de almacenamiento dedicados. Luego almacenan un puntero a ese nodo y los metadatos auxiliares en la cadena de bloques a través de un contrato inteligente. Dependiendo de la forma de este puntero, un usuario puede extraer datos utilizando distintos métodos.

Publish-Subscribe: En este modelo de interacción, los usuarios expresan interés en un elemento de datos (Subscribe) y los dispositivos envían elementos de datos (Publish) a los usuarios interesados. El proceso de suscripción se puede implementar mediante un contrato inteligente, que debe mantener una lista de indicadores para los usuarios interesados. Cada contrato inteligente puede ser responsable de un tema específico y se debe utilizar un mecanismo de resolución para asignar un tema a una dirección determinada de contrato inteligente.

Event-based: Una forma de interacción en el IoT es la actuación (Actuation). Realizar una actuación a través de un contrato inteligente no es trivial, ya que los contratos no interactúan con el mundo físico. Por otro lado, algunas implementaciones de contratos inteligentes proporcionan eventos. Usando eventos, la actuación se puede implementar de la siguiente manera: las operaciones de actuación se pueden implementar como métodos de un contrato inteligente. Los usuarios pueden invocar estos métodos, que a su vez, pueden crear un evento. Los dispositivos podrían monitorear la cadena de bloques en busca de eventos y, si se produce uno, realizarían la actuación adecuada.

3. RESULTADOS OBTENIDOS/ESPERADOS

Como se viene analizando, una de las aplicaciones más importantes de Blockchain que se pueden usar en IoT son los contratos inteligentes.

Los contratos inteligentes son transparentes, se ejecutan de forma determinista por terceros y nadie puede afectar su resultado de ejecución. Proporcionan medios para la autenticación del usuario y la transferencia de tokens. Todas las interacciones con un contrato inteligente se registran en la cadena de bloques. Por otro lado, los contratos inteligentes no son una panacea ya que conllevan riesgos y debilidades. Una vez implementados, no se pueden modificar, no preservan la privacidad del usuario y no pueden almacenar o crear información secreta.

En este trabajo de investigación, se postula que los contratos inteligentes se pueden usar como una abstracción que conectará las aplicaciones con los dispositivos de IoT de una manera eficiente y segura. Unir los dispositivos con contratos inteligentes no siempre será posible, ya que hay dispositivos que no tiene el poder computacional necesario para interactuar con la cadena de bloques.

Además, y teniendo en cuenta que los avances de las investigaciones sobre los protocolos de acceso a los dispositivos y sobre los protocolos de acceso e interoperabilidad de la cadena de bloques avanzan en paralelo, se podría proponer un controlador de protocolos basado en una puerta de enlace (Gateway) que traduciría los protocolos específicos de IoT en transacciones de Blockchain, y viceversa, de manera eficiente y segura.

4. FORMACIÓN DE RECURSOS HUMANOS.

El equipo de investigadores pertenece al cuerpo docente de Tecnologías Aplicadas de la Facultad de Ingeniería, específicamente al área de Seguridad Informática, de la Universidad del Salvador.

A este proyecto, se incorporaron dos docentes investigadores con amplia trayectoria académica, un docente investigador con muchos años de desempeño en la industria de TI y una alumna que se encuentra promediando la carrera de Ingeniería en Informática.

Esto redundará en un aumento del activo académico e investigativo representado por su cuerpo de docentes investigadores, como así también sembrará las bases para la investigación a futuro, a través de la participación de alumnos de la Facultad de Ingeniería.

5. BIBLIOGRAFÍA.

- [1] A. Whitmore, A. Agarwal, and L. Da Xu. “The Internet of Things – A survey of topics and trends”. *Information Systems Frontiers*, vol. 17, nro. 2, pp. 261–274. 2015.
- [2] Glen Martin (Forbes). “How the Internet of Things Is More Like the Industrial Revolution than the Digital Revolution”. <https://www.forbes.com/sites/oreillymedia/2014/02/10/more-1876-than-1995/#674c4e0b66d2>. Última consulta: enero 2022.
- [3] L. Da Xu, W. He, and S. Li. “Internet of things in industries: A survey”. *IEEE Transactions on industrial informatics*, vol. 10, nro. 4, pp. 2233–2243. 2014.
- [4] M. Conoscenti, A. Vetro, and J. C. De Martin. “Blockchain for the internet of things: A systematic literature review”. *Computer Systems and Applications (AICCSA), IEEE/ACS 13th International Conference of. IEEE*, 2016, pp. 1–6. 2016.
- [5] K. Christidis and M. Devetsikiotis. “Blockchains and smart contracts for the internet of things”. *IEEE Access*, vol. 4, pp. 2292–2303. 2016.
- [6] G. Wood, “Ethereum: A secure decentralised generalised transaction ledger,” *Ethereum Project Yellow Paper*, vol. 151, 2014.
- [7] “Hyperledger home page,” *The Linux Foundation*, 2018, (last accessed 23 Feb. 2018). [Online]. Disponible en: <https://www.hyperledger.org/>
- [8] J. Cohn, P. Finn, S. Nair, and P. Sanjai, “Device democracy: Saving the future of the Internet of Things,” *IBM Institute for Business Value*, 2014, Última consulta: enero 2022. Disponible en: [http://www-01.ibm.com/](http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=GBE03620USEN)
- [9] G. C. Polyzos and N. Fotiou, “Blockchain-assisted information distribution for the Internet of Things,” in *Proceedings of the 2017 IEEE International Conference on Information Reuse and Integration*, 2017, pp. 75–78.
- [10] “Smart contracts, ¿Qué son, cómo funcionan y qué aportan?”, Agosto 2016. Disponible en: <https://academy.bit2me.com/que-son-los-smart-contracts/>. Último acceso: Diciembre 2021.
- [11] F. Schüpfer, “Design and Implementation of a Smart Contract Application”, *Master Thesis*, Agosto 2017. Disponible en: <https://files.ifi.uzh.ch/CSG/staff/Rafati/Florian-Schupfer-MA.pdf>
- [12] K. Christidis and M. Devetsikiotis. “Blockchains and smart contracts for the internet of things”. *IEEE Access*, vol. 4, pp. 2292–2303. 2016.

Especificación Integral del Sistema OTP-Vote Orientada a su Implementación

Silvia Bast¹ Germán Montejano² Mario Berón²

¹Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad Nacional de La Pampa
Av. Uruguay 151 – (6300) Santa Rosa – La Pampa – Argentina
Tel.: +54-2954-425166 – Int. 28
silviabast@exactas.unlpam.edu.ar – web: <http://exactas.unlpam.edu.ar/>

²Departamento de Informática
Facultad de Ciencias Físico Matemáticas y Naturales
Universidad Nacional de San Luis
Ejército de los Andes 950 – (5700) San Luis – San Luis – Argentina
Tel.: +54-2652-424027 – Int. 251
[gmonte, mberon]@unsl.edu.ar – web: <http://www.unsl.edu.ar>

RESUMEN

La incorporación del voto electrónico en las sociedades democráticas presenta grandes controversias y discusiones entre los ciudadanos. La mayor resistencia para su implementación pasa por la desconfianza de la sociedad en tales sistemas, debido a las experiencias poco exitosas con las que los usuarios han tenido contacto en diversos lugares en elecciones recientes. Resulta claro que el problema de fondo radica en la confianza de la sociedad sobre el sistema que se usa, por lo que construir sistemas seguros y demostrar la solidez de los mismos, es el principal desafío de investigación de este proyecto.

En 2016 se presentó el modelo inicial de datos de un sistema de voto electrónico denominado OTP-Vote que asegura anonimato incondicional y seguridad computacional que puede llevarse a cualquier nivel exigible. El trabajo expone un modelado básico de los datos de los votos. Para lograr la implementación efectiva del sistema se torna necesario especificar un conjunto de aspectos de gran importancia que quedaron planteados como supuestos en el modelo inicial y que deben aportar las condiciones de seguridad

para lograr un sistema de voto electrónico implementable y robusto. Se presentan a continuación los avances realizados en pos de la especificación integral del sistema.

Palabras clave: *Sistemas de Voto Electrónico, Anonimato, Transparencia, Auditoría, One Time Pad, Verificabilidad End to End.*

CONTEXTO

El presente trabajo tiene sus orígenes en una de las líneas de investigación del proyecto "Aspectos de Seguridad en Proyectos de Software", que avanza en el desarrollo de un modelo de voto electrónico basado en criptografía one time pad. (Resolución N° 488/14 del Consejo Directivo de la Facultad de Ciencias Exactas y Naturales –FCEyN– de la Universidad Nacional de La Pampa - UNLPam).

Del mencionado proyecto surgió una tesis de maestría que presentó las bases del sistema OTP-Vote. Tomando como insumo ese trabajo, se sigue profundizando actualmente sobre la temática desarrollando una tesis doctoral

denominada “Especificación Integral del Sistema OTP-Vote Orientada a su Implementación”, en la Universidad Nacional de San Luis, avalada por Resolución 408/21 de inscripción y Aprobación de Plan de Tesis 408/21 Decanato. Facultad de Ciencias Físico Matemáticas y Naturales -FCFMyN- Universidad Nacional de San Luis - UNSL.

1. INTRODUCCIÓN

Debido a las discusiones y controversias que generan los sistemas de voto electrónico en cuanto a seguridad y transparencia, representan un desafío a los efectos de la investigación. Se torna necesario entonces analizar y evaluar las condiciones de seguridad que deben cumplir y también las soluciones que diferentes autores han propuesto hasta el momento, para intentar generar un modelo teórico que permita luego el desarrollo de un sistema robusto y confiable.

En la tesis de maestría denominada “Optimización de la Integridad de Datos en Sistemas de E-Voting”, defendida en 2016 en la Universidad Nacional de San Luis, se presentaron las bases de un modelo teórico de un sistema de voto electrónico denominado OTP-Vote. En ese trabajo se expone un modelado básico inicial de los datos de los votos, que debe ser refinado, estableciendo precisiones acerca de un conjunto de aspectos, con el fin de lograr una futura implementación.

El refinamiento propuesto apunta a otorgar mayor seguridad a través de:

- Uso de atributos de control y de encriptación, variaciones en cuanto a los datos almacenados de los votos y profundización de las posibilidades de recuperación de los mismos.
- Análisis y refinamiento de protocolos antifraude.
- Análisis y selección de un método criptográfico que asegure la transmisión de datos entre estaciones y servidor.
- Diseño de un modelo para la automatización del proceso de configuración de parámetros y generación de tablas relacionales del sistema de e-voting.

- La información intermedia que puede ser expuesta a los auditores para su control.
- Verificabilidad End to End.

Sistemas de Voto Electrónico

Un sistema de voto electrónico “es un componente de software que mapea electrónicamente el procedimiento de votación” [1].

Importantes autores como Epstein [2], Kazi, Alam y Tamura [3], Prince [4] y van de Graaf, Henrich y Müller-Quade [5], Hao, Ryan [6], Rivest [7], Ryan, Schneide y, Teague [8], Rabin y Rivest [9] y Awad y Leiss [10], expresan detalladamente los requisitos y características de estos sistemas.

El Modelo OTP-Vote

El modelo propuesto en [11] hace uso de i) claves One Time Pad (OTP) que cumplen con la característica de Secreto Perfecto de Shannon [12] (lo que significa que aún un adversario con potencia de cómputo infinita no puede deducir el texto plano a partir del texto cifrado), ii) el esquema de almacenamiento denominado Múltiples Canales Datos único (MCDU) y sus fórmulas propuestas para alcanzar dimensiones con comportamiento óptimo [13], [14], [15] y [16] iii) la operación XOR [17], y iv) la redundancia apropiada [18] para el almacenamiento de los datos para asegurar:

- Anonimato incondicional
- Seguridad computacional que puede llevarse a cualquier nivel exigible durante el proceso electoral.

El proceso de OTP-Vote consiste de tres grandes etapas como puede observarse en la figura 1.

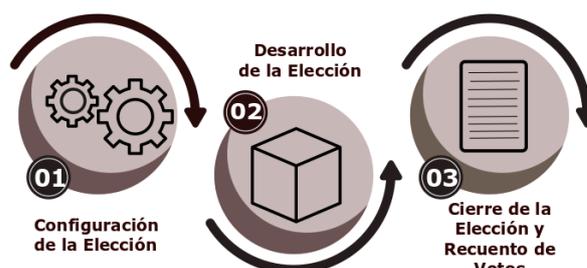


Figura 1. Etapas del Proceso electoral

El modelo teórico presentado supone, para cada una de las etapas mencionadas, el cumplimiento de condiciones de seguridad que resultan imprescindibles para alcanzar el normal funcionamiento del sistema.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Para avanzar en la investigación es necesario realizar un análisis profundo de los aspectos del modelo que requieren de condiciones óptimas de seguridad (comunicación entre usuario-sistema y sistema-servidor de datos).

Una vez identificados los aspectos a mejorar, se realizará una revisión sistemática de aportes de otros autores, para posteriormente trabajar sobre la especificación y validación de propuestas que aporten soluciones en tal sentido.

Será necesario realizar un trabajo minucioso sobre:

- Identificación de los datos que deben permanecer inalterables durante el proceso y los que deben modificarse de forma controlada, para asegurar tales condiciones.
- Análisis de la semántica de las tuplas y especificación propuestas de optimización de configuración que incluyan información de control.
- Especificación y validación de una propuesta de generación automática de tablas relacionales a partir de los datos del sistema.
- Análisis de la información intermedia necesaria para dar transparencia al proceso a la vista de terceros, y especificación, validación y desarrollo de propuestas de auditoría.
- Especificación y validación de una propuesta de verificabilidad End to End.

3. RESULTADOS Y OBJETIVOS

El objetivo de este proyecto es ampliar y refinar la formalización de la especificación del modelo de datos OTP-Vote publicado en 2016, en cuanto a las condiciones de seguridad antifraude para lograr un sistema de voto electrónico robusto, de inmediata y fácil implementación que permita auditorías de control y verificabilidad en todas las etapas de su uso.

Los avances en la investigación están dados por:

- Análisis de los aspectos de seguridad que involucran la comunicación del sistema con el usuario y con el servidor de datos:
 - i) Variaciones en la configuración de los datos en cada proceso eleccionario.
 - ii) Uso de atributos de control y de encriptación.
 - iii) Propuesta inicial de Verificabilidad End to End.
- Revisión sistemática acerca de propuestas ya existentes generadas por otros autores en relación con los aspectos analizados en i), ii) y iii).

Como trabajo futuro, tomando como base los avances ya realizados y la revisión de propuestas ya existentes debe focalizarse en los siguientes aspectos:

- Refinamiento de protocolos antifraude en todas las etapas del modelo.
- Análisis y selección de un método criptográfico que asegure la transmisión de datos entre estaciones y servidor.
- Análisis de la información intermedia que puede ser expuesta a los auditores para su control en las etapas del proceso electoral, sin comprometer el anonimato del votante.
- Elaboración de una propuesta superadora de verificabilidad End to End.
- Evaluación y profundización de los avances ya realizados sobre el modelo original.

4. FORMACIÓN DE RECURSOS HUMANOS

En cuanto a la formación de recursos humanos, Silvia Bast se encuentra desarrollando la tesis denominada “Especificación Integral del Modelo OTP-Vote” para alcanzar el grado de Doctora en Ingeniería Informática en la Universidad Nacional de San Luis, San Luis. Resolución de inscripción y Aprobación de Plan de Tesis 408/21 Decanato. FCFMyN – Universidad Nacional de San Luis.

5. BIBLIOGRAFÍA

[1] B. Ondrisek, B. “E-voting system security optimization”. 42nd Hawaii International Conference on System Sciences (pp. 1-8). IEEE. 2009.

[2] J. Epstein, “Electronic Voting” in *Computer*, vol. 40, no. 8, pp. 92-95, Aug 2007. doi: 10.1109/MC.2007.271.

[3] K. M. Rokibul Alam and S. Tamura, “Electronic voting - Scopes and limitations”, International Conference on Informatics, Electronics & Vision (ICIEV), Dhaka, Bangladesh, pp. 525-529, . 2012. doi: 10.1109/ICIEV.2012.6317324.

[4] A. Prince, “Consideraciones, aportes y experiencias para el Voto electrónico en Argentina”, Editorial Dunken, 2006.

[5] J.van de Graaf, C. Henrich, J. Müller-Quade, “Requirements for secure voting”, Work Notes 2011.

[6] F. Hao, P. Ryan, “Real -World Electronic Voting. Design, Analysis and Deployment”. CRC Press. ISBN-13: 978- 1498714693. ISBN-10: 1498714692. 2017.

[7] R. Rivest, “On the notion of ‘software independence’ in voting systems”.

Philosophical Transactions of the Royal Society A, 366(1881):3759–3767. 2008.

[8] P. Ryan, S. Schneider, V. Teague, “End-to-End Verifiability in Voting Systems, from Theory to Practice”. *Voting Systems, from Theory to Practice. IEEE Security & Privacy*, 13(3):59–62, 2015.

[9] M. Rabin, R. Rivest, “Efficient End to End Verifiable Electronic Voting Employing Split Value Representations” Bregenz, Austria. *Proceedings of EVOTE 2014*. ISBN 978-9949-23-688-6. 2014.

[10] M. Awad, E. Leiss, “End-to-End Cryptography: Spreading Democracy”. *International Journal of Applied Engineering Research*. Volume 11, Issue 11. Ps. 7391-7394. 2016.

[11] S. Bast, “Confidencialidad e Integridad de Datos en Sistemas de E-Voting – Un Modelo para la Implementación Segura de un sistema de Voto Presencial”, Editorial Académica Española. ISBN 978-3-639-53793-2. 2017.

[12] C. E. Shannon, "Communication theory of secrecy systems," in *The Bell System Technical Journal*, vol. 28, no. 4, pp. 656-715, Oct. 1949, doi: 10.1002/j.1538-7305.1949.tb00928.x..

[13] P. García, “Una Optimización para el Protocolo Non Interactive Dining Cryptographers” - Editorial Académica Española (<https://www.eae-publishing.com/> - ISBN-13: 978-3-639-85270-7. ISBN-10: 3639852702. EAN: 9783639852707 – 2017.

[14] J. van de Graaf, G. Montejano, P. García, “Manejo de Colisiones en un Protocolo Non Interactive Dining Cryptographers”. *Anales de las 42° Jornadas Argentinas de Informática e Investigación Operativa (JAIIO, ISSN: 1850-2776)*. Workshop de Seguridad Informática (WSegI 2013, ISSN: 2313-9110). Páginas 29 a 43. 2013 Disponible en: <http://42jaiio.sadio.org.ar/proceedings/simposios/Trabajos/WSegI/03.pdf>.

[15] García P., van de Graaf J., Montejano G., Bast S., Testa O.: “Implementación de Canales Paralelos en un Protocolo Non Interactive Dining Cryptographers”. 43° Jornadas Argentinas de Informática e Investigación Operativa (JAIIO 2014), Workshop de Seguridad Informática (WSegI 2014). <http://sedici.unlp.edu.ar/handle/10915/42066>. 2014.

[16] P. García, J. van de Graaf, A. Hevia, A. Viola, “Beating the Birthday Paradox in Dining Cryptographers Networks”. En “Progress in Cryptology – Latincrypt 2014”. Springer International Publishing. ISSN: 0302-9743. ISSN (electronic): 1611-3349. ISBN: 978-3-319-16294-2. ISBN (eBook): 978-3-319-16295-9. Ps. 179 – 198. Octubre, 2014.

[17] M. Murdocca, V. Heuring, “Principles of Computer Architecture. Appendix A: Digital Logic”. Editor: Addison Wesley; Edición: US ed (29 de noviembre de 1999) Idioma: Inglés - ISBN-10: 0201436647 - ISBN-13: 978-0201436648

[18] P. García, G. Montejano, S. Bast, E. Fritz, "Codificación de Sufragios con Detección de Colisiones en NIDC con Canales Paralelos de Slots” Congreso Nacional de Ingeniería en Informática / Sistemas de Información. CoNaIISI 2016.

Análisis de vulnerabilidades en Videojuegos basados en NFT y criptomonedas

Marco Antonio Villan, mvillan@uade.edu.ar, Departamento de Tecnología, Universidad Argentina de la Empresa, UADE.

Martin Ignacio Melani, mamelani@uade.edu.ar, Departamento de Tecnología, Universidad Argentina de la Empresa, UADE.

1. CONTEXTO

El objetivo de la presente investigación es analizar y evaluar las características de seguridad de los diez juegos con más jugadores basados en NFT Gaming en el periodo 2021/2022.

Se analizarán desde el punto de vista técnico mediante el análisis estático del código de los juegos en búsqueda de posibles vulnerabilidades y un análisis dinámico para analizar los datos que recopilan los videojuegos. Además, se identificarán posibles vulnerabilidades que atenten contra la privacidad y seguridad de los usuarios.

También, se analizarán sus políticas de seguridad, infraestructura, ecosistema NFT y se pondrán en comparación con estándares internacionales en protección de datos y la normativa ISO 27001.

2. INTRODUCCIÓN

En 2021 la capitalización del mercado Non Fungible Token (NFT) alcanzó los 22.000 millones de dólares frente a los 338 millones de dólares de finales de 2020. Esto se debe a que gran parte del sector está centrándose en la industria de los videojuegos (Forbes, 2021).

En el primer trimestre de 2021 los juegos basados en blockchain han pasado de ser juegos del tipo Jugar para Ganar (Play to Earn) a juegos de objetos coleccionables basados en esta modalidad, ya que los NFT raros o poco comunes suelen alcanzar precios astronómicos en el mercado (Binance, 2021).

Actualmente en Argentina hay muchas personas jugando y monetizando. Una persona con una cuenta de Axie Infinity puede más o menos obtener unos 400 dólares por mes, dependiendo del valor de la moneda (SLP) en ese momento (Castro, 2021).

En materia de estadísticas de 2021, los coleccionables de juegos en NFT generaron un total de 2,32 billones en ventas durante el tercer trimestre lo que representa el 22% del volumen total de operaciones de NFT en toda la industria. También, la cantidad de billeteras con la actividad de juego aumento un 2,453% de 29.563 billeteras durante la primera semana de 2021 a 754.000 billeteras en el tercer trimestre de 2021.

En materia de seguridad, los NFT Gaming aseguran la transparencia ya que utilizan la plataforma blockchain para registrar sus operaciones. Además, tiene la propiedad de singularidad y seguridad, los activos del juego no se pueden replicar ni destruir, son únicos y construidos en cadena de bloques (Ray, 2021).

En relación con la seguridad de las NFT, en Understanding Security Issues in the NFT Ecosystem se han detectado problemas de seguridad que tenían gran impacto financiero en los mercados Opensea, Rarible y Sorare. De los marketplaces analizados se detectaron problemas en la autenticación de usuarios, en smart contracts y en token contracts que no son de código abierto y no pueden ser auditados (Das y otros, 2021).

También, existen diferentes escenarios en los que podría comprometerse la seguridad, desde los riesgos en el marketplace, problemas de ciberseguridad, riesgos de fraude, y riesgos en los smart contracts

(Geroni,2021). Se pueden cometer ciberdelitos por otros medios. En 2021, cuentas de usuarios de Nifty Gateway fueron comprometidas y con tarjetas de créditos compraron NFT para luego transferirlos o venderlos (Peters,2021).

3. OBJETIVO GENERAL

Analizar el ecosistema NFT de los juegos más importantes en el periodo 2021/2022 y evaluar sus medidas de seguridad y privacidad. Se analizarán los juegos Aliens World, Axie Infinity, Splinterlands, Bomb Crypto, Sunflower Farmers, Upland, Defi Kingdoms, MOBOX: nft farmer, X worlds games y Elfin Kingdom.

4. OBJETIVOS ESPECÍFICOS

Los objetivos específicos serían los siguientes:

- Analizar si existen vulnerabilidades en los videojuegos mediante el análisis estático y dinámico.
- Identificar el ecosistema NFT
- Identificar vulnerabilidades en el ecosistema NFT
- Definir cuáles son las aplicaciones más utilizadas en Argentina y en Latinoamérica
- Determinar cuáles son las más utilizadas en Argentina.
- Determinar cuál es el cumplimiento de la Ley de Protección de Datos Personales Argentina y del Reglamento General de Protección de Datos de la Unión Europea.
- Determinar qué tipos de datos recopilan los videojuegos
- Establecer la ubicación de las empresas que desarrollan los videojuegos y determinar cuál es la legislación competente en el país.

5. ESPECIFICACIONES METODOLÓGICAS Y/O TÉCNICAS

La propuesta de investigación que se realizará es mixta En cuanto al diseño de investigación será descriptivo, y consisten en analizar la seguridad de los juegos de NFT con más jugadores activos del mercado.

Para la recolección de información se utilizarán las siguientes técnicas:

- Documentales: Revisión y análisis de libros, artículos de revistas, tesis, informes monográficos y páginas web (fichados y análisis de documentos).
- Encuestas de carácter anónimo a personas que utilizan videojuegos para conocer la cantidad de usuarios que la utilizan, sus usos y el juego que utilizan. Se realizará mediante Form office.
- Encuestas para obtener datos estadísticos sobre la cantidad de usuarios promedio y edades en Argentina.
- Entrevistas abiertas y semiestructuradas, las cuales brindan la opción de obtener información
- Entrevistas a especialistas en ciberseguridad
- Desarrolladores de videojuegos
- Especialistas en Criptomonedas}
- Especialistas en NFT
- Especialistas en derecho informático
- Especialistas en privacidad
- Relevamiento y análisis de fuentes secundarias, tales como:
 - Informes de análisis de centros de ciberseguridad
 - Leyes sobre protección de datos personales
 - INCIBE
 - IEEE
 - Google
 - OWASP
 - Common Vulnerabilities and Exposures (CVE)

6. FORMACIÓN DE RRHH

Los participantes de esta investigación (alumnos de grado y maestría) se introducirán al ecosistema NFT y criptomonedas, además, serán capacitados en la utilización de herramientas de análisis de seguridad informática y pentesting.

Los análisis se realizarán en forma automática pero se deberán comparar los resultados con los estándares internacionales en materia de seguridad de la información y privacidad.

7. BIBLIOGRAFÍA

- BINANCE, ¿Qué son los NFT en los juegos? El mercado de los cripto juegos en Blockchain [en línea]. [Fecha de consulta: 12 de enero de 2022]. Disponible en <https://www.binance.com/es/blog/nft/qu%C3%A9-son-los-nft-en-los-juegos-el-mercado-de-los-criptojuegos-en-blockchain-421499824684902192>
- BLOCKCHAIN GAME ALLIANCE. 2021 Member survey & report. [en línea]. [Fecha de consulta: 12 de enero de 2022]. Disponible en <https://www.blockchaingamealliance.org/bga-2021-member-survey-report/>
- CASTRO, Valentina. El Negocio del NFT Gaming. [en línea]. [Fecha de consulta: 12 de enero de 2022]. Disponible en <https://www.pagina12.com.ar/380279-el-negocio-del-nft-gaming>
- DAS, Dipanjan; BOSE, Priyanka; RUARO, Nicola; KRUEGEL, Christopher; VIGNA, Giovanni. Understanding Security Issues in the NFT Ecosystem [en línea]. [Fecha de consulta: 12 de enero de 2022]. Disponible en <https://arxiv.org/abs/2111.08893>
- FORBES DIGITAL. Cripto juegos y juegos NFT: la chance única de acumular dólares sin salir de tu sillón [en línea]. [Fecha de consulta: 12 de enero de 2022]. Disponible en <https://www.forbesargentina.com/money/criptojuegos-juegos-nft-chance-unica-acumular-dolares-salir-tu-sillon-n11000>
- GERONI, Diego. Know The Vulnerabilities And Security Concerns Associated With NFT. [en línea]. [Fecha de consulta: 12 de enero de 2022]. Disponible en <https://101blockchains.com/nft-vulnerabilities-and-security-concerns/>
- PETERS, Jay. Hackers stole NFTs from Nifty Gateway users [en línea]. [Fecha de consulta: 12 de enero de 2022]. Disponible en <https://www.theverge.com/2021/3/15/2231818/nifty-gateway-hack-steal-nfts-credit-card>
- RAY, Allyson. The Impact Of NFTs on Gaming. [en línea]. [Fecha de consulta: 12 de enero de 2022]. Disponible en <https://medium.com/security-token-offering/the-impact-of-nfts-on-the-digital-gaming-domain-b4d0155c03dc>
- STATISTA. NFT games with the highest player count in the last 30 days as of January 10, 2022 [en línea]. [Fecha de consulta: 12 de enero de 2022]. Disponible en <https://www.statista.com/statistics/1266486/blockchain-games-user-number/>
- TURNER, Wright. Edward Snowden dice que los gamers podrían ser vulnerables a exploits mediante los NFT [en línea]. [Fecha de consulta: 12 de enero de 2022]. Disponible en <https://es.cointelegraph.com/news/edward-snowden-says-gamers-could-be-vulnerable-to-exploitation-using-nfts>

Protocolos de consenso

Javier Díaz ¹, Mónica D. Tugnarelli ², Mauro F. Fornaroli ²

Lucas Barboza², Facundo Miño², Juan I. Carubia Grieco²

⁽¹⁾ Facultad de Informática – Universidad Nacional de La Plata

⁽²⁾ Facultad de Ciencias de la Administración – Universidad Nacional de Entre Ríos

e-mail:jdiaz@unlp.edu.ar, monica.tugnarelli, mauro.fornaroli[@.uner.edu.ar]

Resumen

La tecnología blockchain tiene múltiples usos cuando se trata de validar la integridad, la transparencia y trazabilidad de datos. En este trabajo se continúa con la presentación de los avances del PID-UNER 7059 que aborda el estudio de esta tecnología focalizando su aplicación para asegurar la preservación de evidencia digital obtenida de activos esenciales, en un entorno preventivo como lo es Forensic Readiness. Este artículo describe los mecanismos de consenso adoptados por las blockchain seleccionadas así como algunos riesgos de seguridad conocidos y relevantes a la hora de considerar la implementación de una de ellas para cumplimentar el objetivo buscado.

Palabras clave: blockchain, protocolos de consenso, PoA, evidencia digital, seguridad

Contexto

El artículo presenta los avances del Proyecto de Investigación y Desarrollo PID-UNER 7059 denominado “*Tecnología Blockchain para aseguramiento de evidencia digital en entornos Forensic Readiness*” que se encuadra en una de las líneas de investigación establecidas como prioritarias para su fomento, "Arquitectura, Sistemas Operativos y Redes", de la carrera Licenciatura en Sistemas de la Facultad de

Ciencias de la Administración. Se adecua además, a las prioridades de la Universidad Nacional de Entre Ríos por ser un proyecto aplicado a la investigación sobre Tecnologías de la Información y la Comunicación.

Introducción

En artículos anteriores analizamos diferentes características respecto de la tecnología blockchain que incluyeron:

- Relevamiento de casos de uso a nivel regional y nacional, destacando la iniciativa y puesta en funcionamiento de la Blockchain Federal Argentina (BFA) [1].
- Análisis de diferentes tipos de blockchain a los fines de contar con una base de conocimiento previa para la implementación de prototipos en laboratorio, seleccionando para ello dos soluciones representativas de blockchain: una pública basada en Ethereum y otra privada Hyperledger Fabric. Además, se diseñó un esquema genérico del proceso de almacenamiento de hashes que aseguren la evidencia de los activos esenciales determinados. [2]
- Presentación de métricas preliminares sobre la performance de la Blockchain Federal Argentina como ejemplo de Ethereum y una primera revisión del tema sobre Hyperledger Fabric instalada como base de pruebas en laboratorio [3]

En el presente trabajo revisamos los protocolos de consenso utilizados por las soluciones de blockchain seleccionadas para avanzar en la identificación y análisis de riesgos de seguridad que podrían afectarlas.

Es conocido que la tecnología blockchain utiliza criptografía y tecnología de *timestamp* en la capa de datos (Data), que emplea conexiones peer-to-peer para el intercambio de información en la capa de red, que implementa algoritmos de consenso para validar la información y que usa scripts y algoritmos para implementar smart contracts en diversos lenguajes de base. Este entorno de trabajo asegura varias cuestiones relacionadas con la consistencia, la trazabilidad y la inmutabilidad de la información, pero no por ello queda exento a posibles ataques de seguridad que lo comprometan.

El PID 7059 tiene por objetivo primario analizar las prestaciones de la tecnología Blockchain para asegurar la integridad y trazabilidad de la cadena de custodia en un entorno de Forensic Readiness [4], que como método preventivo requiere de respuestas precisas del entorno tecnológico para resguardar los datos considerados como evidencia digital. Por este motivo es imprescindible que esta tecnología asegure un entorno de confianza y privacidad para las partes intervinientes en el posible proceso judicial.

En ese marco, uno de los componentes principales es el mecanismo de consenso utilizado por BFA y por Hyperledger Fabric, tema que se aborda a continuación así como una breve descripción de los principales problemas de seguridad conocidos para dichas implementaciones.

Mecanismos de Consenso

Los mecanismos de consenso, también llamados protocolos o algoritmos de consenso, permiten que los sistemas distribuidos creen un entorno para colaborar y mantenerse seguros. Este mecanismo implica la aceptación de todos los nodos miembros de la blockchain sobre la información que hay en la misma. Es decir, que todos los nodos aceptan que el último bloque ha sido agregado a la cadena de manera correcta, que es el mismo para todos, que no presenta manipulaciones ni datos erróneos.

Tanto BFA como Hyperledger Fabric usan el mecanismo de consenso llamado Proof of authority (PoA) con algunas variaciones según la implementación. [5,6,7]

En la Prueba de Autoridad existen varios nodos de autoridad los cuales están identificados y reciben el nombre de selladores. En este contexto, la identidad significa la identificación personal de un sellador en el mundo real, como por ejemplo en la Blockchain Federal Argentina donde cada nodo sellador debe solicitar su ingreso, presentar documentación legal que certifique su identidad y ser aceptado por $(n/2)+1$ nodos selladores. Como característica distintiva los 21 nodos selladores operativos de la BFA pertenecen a diferentes sectores (académico, industria, organizaciones públicas, privadas).

Esto brinda un control total sobre qué nodos pueden sellar bloques en la red, sirviendo como primera protección para asegurar que un sellador malicioso no pueda generar problemas.

Hyperledger Fabric, de por sí una plataforma permissionada, utiliza el

mecanismo de consenso PoA con base en Kafka Orderer donde los participantes autorizados con acceso controlado validan las transacciones. Cuando la mayoría valida una transacción, hay consenso y se confirma.

Las versiones posteriores de Hyperledger han incorporado el protocolo de consenso Raft basado en el liderazgo, donde los nodos "seguidores" replican las entradas de registro creadas por el "líder" y también pueden elegir un nuevo nodo líder en caso de que este deje de enviar mensajes después de un tiempo configurado. En la red de pruebas se han levantado 5 nodos para simular las partes interesadas.

Además, si bien Kafka es tolerante a fallos, no lo es frente a fallos bizantinos, lo que podría provocar que el sistema no llegue a un acuerdo en el caso de nodos maliciosos o defectuosos.

Entonces, en ambas implementaciones, la identidad de los participantes es conocida abandonando el concepto de anonimato que se asocia a las criptomonedas, lo que sustenta la autenticación de los mismos.

Otro aspecto a considerar es la modalidad de operación, en cuanto a que BFA es descentralizada y todos los nodos pueden acceder al log de registro. En cambio en Hyperledger el registro no es público y tiene un carácter centralizado. Esta última característica tiene ventajas y desventajas conocidas, tales como mayor control de las operaciones, riesgo de que la centralización produzca "cuellos de botella" y que el nodo "líder" sea objeto de un ataque de denegación de servicios, conceptos relacionados con la disponibilidad.

Riesgos de seguridad

En función de lo planteado brevemente en los párrafos anteriores, es necesario relevar las principales vulnerabilidades que afectan a la tecnología blockchain ya que no son inmunes a los ataques y al fraude, tales como los dos que se describen a continuación:

A) Bifurcación (Fork)

Este tipo de incidentes tiene que ver con los cambios en las reglas de consenso, por ejemplo ante una actualización del software de la red de blockchain. Al publicarse una nueva versión del software de la blockchain, cambia el acuerdo sobre las reglas de consenso en los nodos. Las actualizaciones pueden dar lugar a dos tipos de nodos: nodos nuevos, que ya tienen la nueva versión del software, y nodos viejos, que aún no han actualizado a la nueva versión.

Este problema también puede darse por un error tal como sucedió en enero de este año donde la red de pruebas de Ethereum 2.0, Kintsugi, se bifurcó en por lo menos 3 redes diferentes quedando fuera de servicio durante algunos días. En este caso, y según lo informado por el desarrollador de Ethereum [8], el inconveniente se debió a una prueba de verificación de error dentro de la red, que tiene por objetivo crear bloques inválidos cambiando ciertas características, de manera de verificar que los validadores puedan identificar el fallo e invalidar el bloque. Sin embargo, algunos de los nodos selladores identificaron bloques inválidos como correctos, creando una nueva cadena.

Casi al mismo tiempo la BFA sufrió un inconveniente similar donde los selladores quedaron en distintas bifurcaciones

resultando en unas horas sin sellar, lo cual fue solucionado manualmente mediante el grupo de trabajo Selladores levantando y sincronizando los nodos.

Sobre esta problemática en 2020 se realizó una prueba de concepto, conocida como ataque Erebus [9]. Este ataque está específicamente desarrollado para redes P2P como las de blockchain, buscando dividir la red objetivo, mientras que de manera silenciosa se gana el control de la mayor parte de la misma para hacer que efectivamente la red deje de funcionar. Además agrega una manipulación maliciosa en las conexiones con el fin de tomar el control de la misma usando como base un ataque Man-in-the-middle. Como consecuencia la red entra en un estado de no consenso, que termina dividiendo la red y que incluso podría derivar en un ataque de 51% sobre la blockchain.

B) Phishing

En este tipo de ataque se intenta obtener las credenciales de un usuario. Si bien el campo de acción es mayor en las wallets y blockchain con criptomonedas es un riesgo a considerar, mas aun cuando un componente principal para que funcione el entorno seguro que se pretende desplegar es la autenticación de los participantes.

Un posible ataque podría consistir en conseguir los datos de identificación/acceso de un sellador así como otro tipo de información confidencial lo que puede resultar en pérdidas para el usuario, la red blockchain y para el entorno confidencial que se pretende implementar, por lo cual esta vulnerabilidad se deberá considerar tanto para BFA como para Hyperledger Fabric.

Líneas de Investigación, Desarrollo e Innovación

Siguiendo la línea de investigación mencionada en el contexto de este trabajo, se desarrollan actividades que propicien la conformación de una base de conocimiento sobre la tecnología blockchain y sus aplicaciones en diversos ámbitos, destacando el aseguramiento de la integridad y trazabilidad de cualquier activo digital que se considere evidencia digital.

Resultados y Objetivos

El PID 7059 tiene como objetivo primario analizar el impacto de la utilización de la tecnología blockchain aplicada a la preservación, la integridad y trazabilidad de la evidencia digital y la cadena de custodia.

Como objetivos secundarios se establecen:

- Integrar esquemas de recolección de datos y bases de datos de resguardo de evidencia con una solución de blockchain.
- Analizar la relación entre la escalabilidad de blockchain y los algoritmos de consenso.
- Avanzar en la identificación de incidentes de seguridad y el análisis de aspectos de seguridad informática relacionada con la tecnología blockchain.

A la fecha se han cumplimentado diversas etapas en pos del logro de esos objetivos, restando la profundización sobre el tema seguridad el cual es complejo y de permanente actualización.

Por consiguiente, en este artículo se presentaron los resultados preliminares de este análisis, resaltando:

- Si se plantea el resguardo de fragmentos de evidencia forense debe asegurarse un entorno de confianza y privacidad.
- El mecanismo de consenso PoA presenta ventajas tales como eficiencia en los tiempos de transacción y el consenso general de la red, lo que es positivo para la escalabilidad.
- Parte de la ventaja de PoA se convierte también en una vulnerabilidad si consideramos el problema de bifurcación.
- La identificación de los participantes le otorga los requisitos indispensables de transparencia y autenticación, pero se debe tener en cuenta la prevención de una posible suplantación de identidad

Formación de Recursos Humanos

Este proyecto propicia la formación de un docente en co-dirección de proyectos, la formación en actividades de investigación de dos docentes y de un estudiante que se encuentra realizando su trabajo final de la carrera Licenciatura en Sistemas, de un becario del Programa de Becas de Formación (Iniciación en Investigación) de la UNER y de un colaborador estudiante de posgrado de la Maestría en Sistemas de Información que se dicta en la Facultad de Ciencias de la Administración.

Referencias

- [1] Díaz, Francisco Javier; Tugnarelli, Mónica D.; Fornaroli, Mauro F.; Barboza, Lucas. Blockchain para aseguramiento de evidencia digital en entornos Forensic Readiness. XXII Workshop de Investigadores en Ciencias de la Computación (WICC 2020). ISBN: 978-987-3714-82-5
<http://sedici.unlp.edu.ar/handle/10915/103377>
- [2] Díaz, Francisco Javier; Tugnarelli, Mónica D.; Fornaroli, Mauro F.; Barboza, Lucas; Miño, Facundo. Implementación de Blockchain para aseguramiento de evidencia digital en entornos Forensic Readiness. Libro de actas XXVI Congreso Argentino de Ciencias de la Computación (CACIC 2020). ISBN: 978-987-4417-90-9
<http://sedici.unlp.edu.ar/handle/10915/113243>
- [3] Díaz, Francisco Javier; Tugnarelli, Mónica D.; Fornaroli, Mauro F.; Barboza, Lucas; Miño, Facundo. Métricas para blockchain. Libro de actas XXVII Congreso Argentino de Ciencias de la Computación (CACIC 2021). ISBN: 978-987-633-574-4
<http://sedici.unlp.edu.ar/handle/10915/129809>
- [4] Tan, John. (2001). Forensic Readiness. http://isis.poly.edu/kulesh/forensics/forensic_readiness.pdf
- [5] Mecanismos de Consenso Ethereum. <https://ethereum.org/es/developers/docs/consensus-mechanisms/>
- [6] Mecanismos de Consenso BFA. <https://bfa.ar/blockchain/protocolos-de-consenso>
- [7] Mecanismos de Consenso Hyperledger Fabric https://hyperledger-fabric.readthedocs.io/es/latest/fabric_model.html
- [8] Marius van der Wijden (2022) <https://twitter.com/vdWijden/status/1479377978400419843>
- [9] Muoi Tran, Inho Choi, Gi Jun Moon, Anh V. Vu, Min Suk Kang. A Stealthier Partitioning Attack against Bitcoin Peer-to-Peer Network . 2020 IEEE Symposium on Security and Privacy <https://erebus-attack.comp.nus.edu.sg/erebus-attack.pdf>
- [10] Iuon-Chang Lin, Tzu-Chun Liao. A Survey of Blockchain Security Issues and Challenges. <https://pdfs.semanticscholar.org/f61e/db500c023c4c4ef665bd7ed2423170773340.pdf>
- [11] H. Chen, M. Pendleton, L. Njilla, and S. Xu. 2020. A Survey on Ethereum Systems Security: Vulnerabilities, Attacks, and Defenses. ACM Computing Surveys. 53, 3, Article 67 (June 2020)

Criptografía liviana para objetos conectados

Universidad Argentina de la Empresa

Ricardo Wehbe – rwehbe@uade.edu.ar

Adrián de Armas – adearmas@uade.edu.ar

Elizabeth Barrera – elbarrera@uade.edu.ar

1 Plan de investigación

1.1 Contexto

En los próximos años se acentuará una tendencia que ya ha comenzado, la interconexión de diversos objetos “inteligentes”, denominada IoT (Internet of Things) [1]. Esto trae aparejado un problema de seguridad: los métodos habituales de criptografía simétrica y asimétrica (AES, DES, block ciphers entre los primeros; RSA, Diffie-Hellman, ElGamal, curvas elípticas entre los segundos) [2] [3] requieren un esfuerzo computacional del que estos dispositivos son incapaces.

La criptografía liviana (lightweight cryptography) [4] [5] tiene por objetivo lograr un aceptable nivel de seguridad para dispositivos de capacidades modestas.

Este proyecto se propone evaluar comparativamente los métodos de criptografía liviana propuestos actualmente desde el punto de vista de su consumo energético y los recursos computacionales requeridos.

1.2 Estado actual de conocimiento sobre el tema

Existen diversos algoritmos de criptografía liviana actualmente. Se los puede dividir en:

1. *Stream ciphers* (algoritmos que encriptan un flujo continuo de datos) como SNOW 3G, Trivium, Espresso, LIZARD, etc.
2. Funciones de *hashing* criptográficas (si bien no se utilizan directamente para encriptar, son un componente esencial de muchos esquemas de seguridad). Algunos ejemplos son Armadillo, QUARK, Photon, GLUON, etc.
3. *Block ciphers* (algoritmos que encriptan un flujo de datos por bloque) como TEA, RC5, Hight, PRESENT, KATAN, etc.
4. Algoritmos de criptografía asimétrica, (sistemas de clave pública), como RSA y curvas elípticas (ECC).

Dos buenos *surveys* son los de Bhardwaj et alii y Biryukov et alii [6] y [7], aunque el primero es excesivamente lacónico y el segundo sólo considera criptografía simétrica.

En este proyecto se pretende incluir a los principales algoritmos de criptografía liviana asimétrica, como RSA o ECC [8] [9]. No se ha encontrado antecedentes de estudios comparativos como el que querríamos realizar en esta área. En una etapa ulterior, queremos trabajar sobre funciones de *hashing* livianas y stream ciphers.

2 Objetivos

2.1 Objetivos generales

El objetivo del proyecto es efectuar un relevamiento de los algoritmos actuales de criptografía liviana, implementar estos algoritmos y compararlos desde el punto de su consumo energético y costo computacional.

2.2 Objetivos específicos

La idea es tener un panorama general de las ventajas y desventajas de los diferentes abordajes y de adquirir un *know-how* en la implementación y análisis de estos algoritmos.

Las diferentes familias de algoritmos se considerarán por separado (algoritmos simétricos, asimétricos y funciones de *hashing* [10].) Se implementarán estos algoritmos y se comparará la eficiencia relativa de cada uno de ellos para las diferentes familias consideradas. En una primera etapa nos conformaremos con implementaciones en software, aunque el objetivo final es llegar a las implementaciones en hardware.

3 Especificaciones metodológicas y técnicas

En una primera etapa se hará una selección de algoritmos en tres categorías (criptografía liviana simétrica, criptografía liviana asimétrica y funciones de hashing livianas).

Para cada una de las categorías se efectuará una implementación lo más eficiente posible, utilizando las técnicas que se ven en las materias

de programación avanzadas [11] [12]. Existen además de las técnicas generales, algunas técnicas específicas para este género de problemas (por ejemplo, la utilización de primitivas de las versiones livianas de AES para producir funciones de *hashing* livianas [13].)

El análisis procederá por dos caminos paralelos. Por una parte, estimar el orden de complejidad temporal a través del análisis del código con algunas de las notaciones asintóticas estándar (la notación O o θ [14][15].) Esto dará una estimación general de la eficiencia de un algoritmo. Pero, como se sabe, estas notaciones dejan de lado constantes que pueden ser importantes.

Por eso, se establecerá una medida de la duración real de ejecución sobre un conjunto de datos establecido para determinar la posible influencia de las constantes que las notaciones asintóticas dejan de lado. Se busca crear un ambiente de prueba en el que todos los algoritmos estén trabajando en igualdad de condiciones para que los resultados no estén sesgados.

Para la segunda etapa, la confidencialidad se conseguirá a través de un esquema de criptografía simétrica. Para la autenticación de datos de origen se utilizará, como es habitual, un esquema de clave privada. Se buscará establecer un esquema de confianza en la red para que los objetos estén en condiciones de aceptar o rechazar pedidos de conexión (un esquema similar al que se utiliza en [27] para redes de sensores.) Para ello se asume un contexto de ataques bizantinos [25], en el que se trata de mantener la coherencia de la información en un contexto en el que algunos nodos de la red pueden estar transmitiendo información poco confiable [26].

Un importante objetivo es la posibilidad de crear un grupo de interés que pueda eventualmente expandirse en otras direcciones y estar en el mediano plazo en condiciones de efectuar transferencia tecnológica.

Cabe destacar que algunos de los objetivos ya están cumplidos. Se ha realizado una comparación de diversos algoritmos de clave pública, clave simétrica y funciones de *hashing* livianas en tres diferentes trabajos finales de ingeniería. En todos los casos, se trató de implementaciones en software.

4 Formación de recursos humanos

Los participantes de esta investigación (alumnos de grado y maestría) serán capacitados para entender los fundamentos matemáticos de las principales herramientas criptográficas (funciones de *hashing* criptográficas, algoritmos de criptografía simétrica y asimétrica, MAC, firma digital, métodos de autenticación. Se imple-

mentarán algoritmos seleccionados en distintos lenguajes de programación y se realizará la verificación de los algoritmos implementados y la comparación entre las distintas implementaciones para determinar puntos de referencia respecto del mejor algoritmo para cada escenario.

En una etapa ulterior se prevé trabajar sobre hardware (posiblemente Raspberry.) Para ello, se planea un trabajo conjunto con la Facultad de Ingeniería.

5 Bibliografía

- [1] Li, Shancang; Xu, Li Da; Zhao, Shanshan: The Internet of Things: a Survey, *Inf. Syst. Front.* 17, pp. 243–259, 2015.
- [2] Delfs, Hans; Knebel, Helmut: *Introduction to Cryptography*, Springer, 2007.
- [3] Cohen, Henri; Frey, Gerhard (eds.), *Handbook of Elliptic and Hyperelliptic Cryptography*, Chapman & Hall, 2007.
- [4] Buchanan, William; Li, Shancang; Assiz, Rameez: *Lightweight Cryptography Methods*, *Journal of Cyber Security Technology*, 1(3-4), pp. 187–201, 2017.
- [5] Dutta, Indira; Bayoumi, Magdy; Ghosh, Baskar: *Lightweight Cryptography for Internet of Insecure Things: a Survey*, Proc. of the 9th Annual IEEE Computing and Communication Workshop and Conf. (CCWC), 2019.
- [6] Bhardwaj, Isha; Kumar, Ajay; Bansal, Manu: *A Review on Lightweight Cryptography Algorithms for Data Security and Authentication in IoT*, Proc. of the 4th IEEE Int. Conf. On Signal Processing, Computing and Control (IC-SPCC), 2017.
- [7] Biryukov, Alex; Perrin, Léo-Paul: *State of the Art in Lightweight Symmetric Cryptography*, *IACR Cryptology ePrint Archive* 2017: 511, 2017.
- [8] Abd Zaid, Mustafa ; Hassan, Soukaena : *A Lightweight RSA Algorithm Using Three Prime Numbers*, *Int. J. of Engineering & Technology*, 7(4.36), pp. 293–295, 2018.
- [9] Lara-Nino, Carlos; Díaz-Pérez, Arturo; Morales-Sandoval, Miguel: *Elliptic Curve Lightweight Cryptography: a Survey*, *IEEE Access*, 6, pp. 72514–72550, 2018.
- [10] Hammad, Baraa Tareq; Jamil, Norziana; Rusli, Mohd Ezanee; Z'aba, Muhammad Reza: *A survey of Lightweight Cryptographic Hash Functions*, *Int. J. of Scientific & Engineering Research* 8(7), pp. 806–814, 2017.
- [11] Cormen, Thomas; Leiserson, Charles; Rivest, Ronald; Stein, Clifford: *Introduction to Algorithms*, MIT Press, 2002.
- [12] Dasgupta, Sanjoy; Papadimitriou, Christos; Varizani, Umesh: *Algorithms*, McGraw-Hill Education, 2006.

- [13] Bos, Joppe; Özen, Onun; Stam, Martijn: *Efficient Hashing Using the AES Instruction Set*, Proc. of the Conf. on Cryptographic Hardware and Embedded Systems (CHES) 2011, pp. 507–522, 2011.
- [14] Lewis, Harry; Papadimitriou, Christos: *Elements of the Theory of Computation*, Prentice-Hall, 1998.
- [15] Sipser, Michael: *Introduction to the Theory of Computation*, Thomson Course Technology, 2006.
- [16] Cheruvu, Sunil; Kumar, Anil; Smith, Neil; Wheeler, David: *Demystifying Internet of Things Security*, Apress Open, Springer, 2020.
- [17] Hou, Jianwei; Qu, Leilei; Shi, Wenchang: *A survey on internet of things security from data perspectives*, Computer Networks 148(15), pp. 295–306, 2019.
- [18] Veltri, Giuseppe: *Digital Social Research*, Polity Press, 2020.
- [19] Stephens-Davidowicz, Seth: *Everybody Lies. Big Data, New Data, and What the Internet Can Tell Us about Who We Really Are*, Bloomsbury Publishing, 2017.
- [20] Ohlhorst, Frank: *Big Data Analytics. Turning Big Data Into Big Money*, John Wiley & Sons, 2013.
- [21] Alexandre, Laurent: *La guerre des intelligences : intelligence artificielle versus intelligence humaine*, JC Lattès, 2017.
- [22] Liu, Donggang; Ning, Peng: *Security for Wireless Sensor Networks*, Springer, 2007.
- [23] Jurdak, Raja: *Wireless Ad Hoc and Sensor Networks. A Cross-Layer Design Perspective*, Springer, 2007.
- [24] Sohraby, Kazem; Minoli, Daniel; Znati, Taieb: *Wireless Sensor Networks. Technology, Protocols and Applications*, John Wiley & Sons, 2007.
- [25] Lamport, Leslie: *The Byzantine Generals Problem*, ACM Trans. on Programming Languages and Systems 4(3), pp 387-389, 1982.
- [26] Abdelhakim, Mai; Lightfoot, Leonard; Li, Tontong: *Reliable Data Fusion in Wireless Sensor Networks under Byzantine Attacks*, Proc. of the 2011 IEEE Military Comm. Conf. (MILCOM 2011), pp. 810-815.
- [27] Chen, Haiguang; Wu, Huafeng; Zhou, Xi; Gao, Chuanshan: *Agent-Based Trust Model in a Wireless Sensor Network*, Proc. of the 8th IEEE Conf. on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (ACIS), pp. 119-124, 2007.

Ciberdefensa y los Usos Maliciosos de la Criptografía.

Cipriano, Marcelo^{1,2}; García, Edith¹, Maiorano, Ariel¹
Malvacio, Eduardo¹, Pazo Robles, María Eugenia¹

¹Laboratorio de Informática, Software Seguro y Criptografía, Facultad de Ingeniería del Ejército (FIE),
Universidad de la Defensa Nacional - UNDEF

²Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes UNQ.

{marcelocipriano; egarcia; maiorano; emalvacio; mepazorobles}@fie.undef.edu.ar

RESUMEN

Los objetivos principales del proyecto son el estudio, análisis, paradigmas y herramientas criptológicas modernas destinadas a la creación de software malicioso y puertas traseras criptográficas. Asimismo indagar mecanismos y metodologías que posibiliten la prevención, detección y protección para ser aplicadas en el ámbito de la Ciberdefensa Nacional.

La criptografía y todas sus aplicaciones (firma digital, autenticación, etc.) suele ser conocidas por ser de las más importantes herramientas de carácter defensivo. Los últimos años esa idea viene levemente cambiando hacia el sentido opuesto. Es decir que puede emplearse en aplicaciones ofensivas y maliciosas. Las más difundidas, al menos hasta ahora, son el secuestro, la pérdida de información y la extorsión. El responsable de llevar adelante tales actos, es una nueva clase de malware, conocida como *Ransomware*.

Otra aplicación, no tan conocida y de igual o mayor impacto aún, es el diseño e implementación de algoritmos criptográficos que incluyan las llamadas *Backdoors Cryptography* o puertas traseras criptográficas. Estas pueden vulnerar la confidencialidad, integridad y disponibilidad de la información que, en teoría, deberían proteger.

Son evidentes las graves consecuencias sobre la ciberseguridad de usuarios particulares, empresas y organismos no gubernamentales de tales ataques.

Mayor aún es el impacto que podrían tener sobre la protección de las Infraestructuras Críticas (IICC) de una sociedad, los sistemas de información y comunicación empleados en las fuerzas que defienden la nación, como así también la amenaza directa sobre la población.

Si las IICC están comprometidas, en forma directa o indirecta, entonces la vulnerabilidad trasciende a la ciberseguridad, sino que afecta directamente a la Ciberdefensa. Entendidas aquí a las IICC como organizaciones relacionadas con la generación y distribución de energía, sistema financiero y bancario, organismos de salud como hospitales, servicio de potabilización y distribución de agua, saneamiento de desechos, entre otras.

Palabras Clave

Criptología, Criptovirología, Kleptografía, Puertas Traseras Criptográficas. Ciberdefensa.

CONTEXTO

El proyecto “MAC: Criptografía Maliciosa para la Ciberdefensa” pertenece a la Facultad de

Ingeniería del Ejército (FIE) "Gral. Div. Manuel N. Savio", perteneciente a la Universidad de la Defensa Nacional (UNDEF).

Se enmarca en el contexto de la carrera de grado de Ingeniería en Informática, la Especialización en Criptografía y Seguridad Teleinformática (Ciberseguridad) y la Maestría en Ciberdefensa, que se dictan en la citada unidad académica.

Los investigadores conforman el *Grupo de Investigación en Criptología y Seguridad Informática (GICSI)* que pertenece al *Laboratorio de Informática, Software Seguro y Criptografía* y lleva adelante tareas de *I+D+i*.

El equipo está conformado por docentes investigadores categorizados en distintos regímenes científicos, profesionales técnicos y alumnos de las carreras antes mencionadas.

1. INTRODUCCIÓN

Suele reconocerse como instrumentos defensivos a la Criptografía cuando ofrece en sistemas de comunicaciones, redes y bases de datos -entre otros- confidencialidad, integridad y autenticación.

Pero esa visión demostró ser parcial, pues en 1996 *Adam Young* y *Moti Yung* [1] presentan lo que han dado en llamar *Criptovirología*.

Ellos demostraron la posibilidad convertir la criptografía en una herramienta de ataque. Asociada adecuadamente a un virus informático, se convierte en un poderoso vector de ataque. Primeramente se infecta el sistema mediante el virus informático. Una vez dentro del sistema, el virus activa su *payload* malicioso, cifrando la información de su víctima. Finalmente, muestra por pantalla un mensaje extorsivo, donde le pide al usuario que

pague rescate si quiere recuperar la información así "secuestrada".

Lo que los autores estaban describiendo en aquel trabajo fundacional, era lo que hoy se conoce como *ransomware*¹ considerado un malware de alto impacto y profunda afectación de los sistemas que ataca.

En 1997, los mismos autores presentan la llamada "*Kleptografía*": esto es el diseño e implementación de *Backdoors Cryptography* o *Puertas Traseras en Algoritmos Criptográficos* [2-4].

Para demostrar la viabilidad de la *Kleptografía* más allá de su concepción abstracta, los autores presentan a "*Secretly Embedded Trapdoor with Universal Protection*", *SETUP*.

Este *kleptograma* es una modificación a nivel matemático del algoritmo de intercambio de claves *Diffie-Hellman (DH)*, considerada la mejor herramienta para resolver el problema de la *Distribución de Claves* que durante milenios estuvo sin solución.

¿Qué consecuencias acarrearían la aplicación de tales *técnicas kleptográficas* en otros algoritmos? Por mencionar algunos ejemplos: esquemas de cifrado y de firma digital *ElGamal*, *DSA*, el algoritmo de firma de *Schnorr*, y el *PKCS* de *Menezes-Vanstone* y finalmente el reconocido algoritmo *RSA* [5-6, 8,13], entre otros.

Se podría asumir que la *Keptografía* se acota a la *Criptografía Asimétrica o de Clave Pública*. Sin embargo se encuentran publicadas funciones hash. Por ejemplo una versión de *SHA-1* modificado[7].

También se pueden hallar alternativas para protección de funciones hash comprometidas en algoritmos de nivel superior, como *HMAC* y *HKDF* [14].

¹ Ransom: rescate en inglés.

Tampoco quedan indemnes los generadores de números pseudo-aleatorios, conocidos como *Pseudo Random Numbers Generators* o *PRNG* [10-13]. A rasgos generales, se inserta una vulnerabilidad en el núcleo de la primitiva criptográfica, que afecta las propiedades estadísticas de tales generadores.

La seguridad de los esquemas criptográficos se mide tradicionalmente, entre otras consideraciones, como la incapacidad de un adversario de violar un objetivo de seguridad deseado, al contar con ciertas limitaciones en sus recursos [14].

Se observa que este argumento de seguridad se basa en un diseño sólido de los componentes subyacentes a los que un adversario no tiene acceso para afectar o influir. Si estas consignas no se satisfacen, es decir que un adversario pueda influir en el diseño de un algoritmo aplicando *Kleptografía*, entonces las consecuencias devastadoras de tal acción, resultan evidentes[16,17].

Peor aún, el daño que podría ocasionar si tal algoritmo, además, fuera estandarizado mediante alguna norma y/o adoptado por algún protocolo o aplicación.

Nadie puede aseverar que este escenario aquí propuesto es sólo una presunción realizable a futuro, o una realidad presente.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

El proyecto sigue distintas líneas de acción:

- Estudio de material actualizado, asistencia a Cursos, Congresos y Workshops específicos, profundización en el estado del arte

tanto de la *Criptovirología* como de la *Kleptografía*, aunque el esfuerzo principal estará dirigido a esta última.

- Estudio y análisis de las diferentes variantes de *Criptovirología*.

- Estudio y análisis de ataques *kleptografía* en la literatura aplicados a diferentes algoritmos o primitivas criptográficas.

- Profundización en el estudio y análisis de técnicas *kleptográficas* para el algoritmo *RSA* y algoritmos de *generación de números pseudo-aleatorios* específicamente.

- Implementación experimental, conceptual y de referencia de alguna o algunas de las técnicas analizadas.

- Análisis y conclusiones de los resultados obtenidos.

3. RESULTADOS OBTENIDOS / ESPERADOS

El proyecto persigue estudiar y analizar la aplicación de los paradigmas y herramientas en la creación de software malicioso y puertas traseras criptográficas, para así poder desarrollar técnicas de prevención, detección y protección para ser considerados en el ámbito de la Ciberdefensa Nacional.

4. FORMACIÓN DE RECURSOS HUMANOS

Los docentes del equipo de investigación dictan distintas asignaturas en las carreras de grado y posgrado en FIE. Y desde dichas cátedras se invita de forma permanente a los alumnos para participar como colaboradores.

Se han incorporado al equipo algunos alumnos de la especialización en *Criptografía* y *Seguridad*

Teleinformática (Ciberseguridad) que están llevando a cabo su *Trabajo Final Integrador (TFI)* como así también alumnos de la *Maestría en Ciberdefensa*, que se encuentran trabajando en el desarrollo de sus respectivas tesis. En ambos casos se abordan temáticas afines a la de este proyecto.

Se espera que la contribución mutua entre el equipo de investigadores, especializandos y maestrandos permita alcanzar niveles sinérgicos de avance en la investigación, la formación de recursos humanos.

La Formación de Recursos Humanos permite incrementar el Know-How que tendrá el grupo de investigadores a lo largo de la vida del proyecto. Será un importante beneficio de sus integrantes y de la institución en la cual desarrollan sus actividades científico-docentes.

Por último y atendiendo a la responsabilidad ética y social que compete a la actividad científica y tecnológica, el Grupo Integrante de este Proyecto de Investigación, ya sea durante su ejecución o por la aplicación de los resultados obtenidos, desea expresar su compromiso a no realizar cualquier actividad personal o colectiva que pudiera afectar los derechos humanos, o ser causa de un eventual daño al medio ambiente, a los animales y/o a las generaciones futuras.

5. BIBLIOGRAFÍA

- [1] Young, Adam L. and Moti Yung. "Cryptovirology: extortion-based security threats and countermeasures." Proceedings 1996 IEEE Symposium on Security and Privacy (1996): 129-140.
- [2] Young, Adam L. and Moti Yung. "The Prevalence of Kleptographic Attacks on Discrete-Log Based Cryptosystems." CRYPTO (1997).
- [3] Young, Adam L. and Moti Yung. "Kleptography: Using Cryptography Against Cryptography." EUROCRYPT (1997).
- [4] Young, Adam L. and Moti Yung. "Malicious cryptography - exposing cryptovirology." (2004).
- [5] Young, Adam L. and Moti Yung. "A Space Efficient Backdoor in RSA and Its Applications." Selected Areas in Cryptography (2005).
- [6] Young, Adam L. and Moti Yung. "An Elliptic Curve Backdoor Algorithm for RSASSA." Information Hiding (2006).
- [7] Albertini, Ange, Jean-Philippe Aumasson, Maria Eichlseder, Florian Mendel and Martin Schl affer. "Malicious Hashing: Eve's Variant of SHA-1." Selected Areas in Cryptography (2014).
- [8] Young, Adam L. and Moti Yung. "Cryptography as an Attack Technology: Proving the RSA/Factoring Kleptographic Attack." The New Codebreakers (2015).
- [9] Russell, Alexander, Qiang Tang, Moti Yung and Hong-Sheng Zhou. "Cliptography: Clipping the Power of Kleptographic Attacks." ASIACRYPT (2015).
- [10] Indarjani, Santi. Sugeng, Kiki. Widjaja, Belawati. "Modification Attack Effects on PRNGs: Empirical Studies and Theoretical Proofs." (2015).
- [11] Young, Adam L. and Moti Yung. "Cryptovirology: the birth, neglect, and explosion of ransomware" Commun. ACM 60 (2017): 24-26.
- [12] Teseleanu, George. "Random Number Generators Can Be Fooled to Behave Badly." IACR Cryptology ePrint Archive (2018).
- [13] Markelova, A. V. "Vulnerability of RSA Algorithm." (2018).
- [14] Fischlin, Marc. Janson, Christian. Mazaheri, Sogol. "Backdoored Hash Functions: Immunizing HMAC and HKDF." (2018): 105-118.
- [15] Xiao, Dianyan and Yang Yu. "Klepto for Ring-LWE Encryption." Comput. J. 61 (2018): 1228-1239.
- [16] Yogi, Manas. Aparna, S.. "Novel insights into Cryptovirology A Comprehensive Study." International Journal of Computer Sciences and Engineering. 6. (2018): 1252-1255.
- [17] Zimba, Aaron. Chishimba, Mumbi. "On the Economic Impact of Crypto-ransomware Attacks: The State of the Art on Enterprise Systems." European Journal for Security Research. (2019).

Criptografía Liviana para aplicar en IoT e IIoT

Cipriano, Marcelo; Eterovic, Jorge; García, Edith; Torres, Luis.

Instituto de Investigación en Ciencia y Tecnología
Dirección de Investigación Vicerrectorado de Investigación y Desarrollo.
Universidad del Salvador.
Lavalle 1854 – C1051AAB -Ciudad Autónoma de Buenos Aires - Argentina

{ marcelo.cipriano; jorge.eterovic; edith.garcia }@usal.edu.ar
luis.antonio.torres@kyndryl.com

RESUMEN

Desde hace ya varios años existen algoritmos criptográficos que por sus características operativas, conforman la llamada Criptografía Liviana o Ligera. Tales algoritmos pueden emplearse en dispositivos con bajos recursos, como son los que pertenecen a la Internet de las Cosas (IoT) y la Internet Industrial de las Cosas (IIoT), también llamada Industria 4.0. Este proyecto persigue el estudio y análisis de tales algoritmos y los protocolos que los utilizan.

Además del abordaje matemático y algorítmico de la temática a investigar, el proyecto también se propone metas de Difusión y Transferencia de estas temáticas. No solamente orientadas al ámbito académico de las Tecnologías de la Información, también conocido como IT (Information Technology por sus siglas en inglés). Sino también y en la medida de lo posible, se alcance la difusión en el ámbito de la producción industrial y las empresas dedicadas a ello, donde las Tecnologías de las Operaciones u OT (Operation Technology) desarrollan sus actividades. Corazón de la llamada “Cuarta Revolución Industrial”, cada vez más interconectada y con requerimientos de seguridad en aumento.

El proyecto desea contribuir con ambos mundos tecnológicos, cada vez más cercanos, cuyo horizonte a mediano plazo será seguramente, la convergencia en los llamados “Sistemas Ciber-Físicos”.

La Criptografía Liviana [1] permite dotar de Confidencialidad, Integridad y Autenticación a dispositivos IoT, en los que la Criptografía convencional no puede aplicarse, por sus elevados requerimientos de recursos de procesamiento, cálculo, memoria, energía y demás.

Por otro lado, muchos fabricantes por razones que valdría la pena profundizar, pero que se escapan a los alcances del proyecto, no dotan de mecanismos de seguridad a sus diseños, dispositivos y equipos. Es un gran desafío a corto plazo, subsanar estas falencias que podrían poner el serio riesgo el mundo OT.

Palabras Clave:

Criptografía Ligera, Internet de las Cosas, Internet de las Cosas Industrial, IoT, IIoT.

CONTEXTO

El Vicerrectorado de Investigación y Desarrollo (VRID), perteneciente a la Universidad Nacional del Salvador (USAL), dicta las políticas referidas a la investigación, concibiéndolas como un servicio a la comunidad, entendiendo que los nuevos conocimientos son la base de los cambios sociales y productivos. Con el impulso de las propias Unidades Académicas se han venido desarrollando acciones conducentes a concretar proyectos de investigación uni/multidisciplinarios, asociándose a la docencia de grado y postgrado y vinculando este accionar, para potenciarlo, con otras instituciones académicas del ámbito nacional e internacional.

La Dirección de Investigación, dependiente del VRID, brinda soporte a las distintas Unidades de Investigación y a sus investigadores para el desarrollo de Proyectos y Programas de Investigación, nacionales e internacionales, como así también, apoyo y orientación de recursos para la investigación.

A ella pertenece el Instituto de Investigación en Ciencia y Tecnología (RR 576/12) en el cual se

enmarca este proyecto con una duración de 2 años (2021-2023).

1. INTRODUCCIÓN

Las llamadas “*Smart Cities*” o *Ciudades Inteligentes* son aquellas que adoptan sensores y dispositivos IoT para llevar adelante tareas de gestión, control y servicios que le son propias. Sin embargo, la novedad es que adicionalmente a las funcionalidades que de ellos se esperan, también se obtiene, entre otros:

- Uso racional de los recursos.
- Ahorro energético (importante factor a tener en cuenta en el contexto del cuidado del medio ambiente, la reducción de emisiones de CO₂ y el cambio climático, claros desafíos de la humanidad para el siglo XXII).
- Reducción en los tiempos de mantenimiento.
- Control a distancia e interconexión mediante una red de datos (inalámbrica usualmente).
- Comunicación interactiva entre sí y con el Centro de Comando y Control (C2).
- Información en tiempo real de los datos recabados y estado de funcionamiento.

Distintas ciudades argentinas han ingresado a la clase de las *Smart Cities*, desde hace ya varios años. Ellas han adoptado diferentes sistemas y equipos IoT. Por ejemplo, desde 2015 calles de la Ciudad Autónoma de Buenos Aires cuentan con un *Sistema de Telegestión* [1] que puede controlar individualmente el alumbrado público, a través de las llamadas “*luminarias inteligentes*” y desde 2016 las autopistas de la ciudad [2] cuentan con este tipo de iluminación, aprovechando todas las ventajas que de ellas se obtiene tal como fueron mencionadas anteriormente.

Los aportes no se agotan en las luminarias. Sensores de contaminación, ruido y tránsito, cámaras de seguridad con reconocimiento facial y lectura de patentes extienden las aplicaciones de IoT en nuestras ciudades, como Salta, Mendoza y Bahía Blanca [3] entre otras.

Pero por otro lado, estos y otros dispositivos IoT e IIoT presentan vulnerabilidades susceptibles de ser explotadas. Un aporte sustancial a la seguridad, entre otros, es la aplicación de confidencialidad y/o autenticación mediante *Criptografía Liviana o Ligera* [4-5].

Un análisis superficial del problema podría llevar a concluir que la inseguridad queda encerrada en

el pequeño ámbito de las luminarias o los sensores antes mencionados. Y por lo tanto, es fácilmente de resolver. Por el contrario, el crecimiento exponencial de la cantidad de dispositivos IoT que se suman al ecosistema informático [6-7] y contribuye a aumentar el riesgo

El mundo IoT sorprende a los expertos en seguridad: la mayoría de tales dispositivos *no presentan mecanismos de seguridad o los mismos son rudimentarios*. Este grave problema ya fue expuesto en forma extensa [8-10], conjuntamente al reclamo por su pronta resolución.

La información de los usuarios se puede ver comprometida severamente, como así también el dispositivo en sí mismo.

Es evidente el nivel de gravedad de esta situación. Pero aún no se ha visto el panorama completo. Está en ciernes un aspecto más grave aún, un punto débil que ya ha sido explotado y nada impediría que se repita en un futuro cercano. O peor aún, que escale su efecto nocivo.

Los dispositivos IoT e IIoT amplían (incluso a un ritmo exponencial) la llamada “*Superficie de Ataque*”. Esta situación pone en riesgo a la propia interconexión mundial de las redes. Y ya ocurrió pues el 21 de Octubre de 2016, millones de dispositivos en todo el planeta, la mayoría de ellos IoT, produjeron un ataque *Denegación de Servicio Distribuido (DDoS)* direccionado contra un proveedor de servicio de *DNS*. Este ataque afectó un servicio crítico de la propia Internet. No fue el primero de tu tipo, pero si el mayor registrado hasta el presente.

Este exitoso ataque afectó a empresas como *Amazon, BBC, CNN, Fox News, Github, HBO, Netflix, New York Times, PayPal, Spotify, Starbucks, Twitter, Visa, Wall Street Journal*, entre otras.

El responsable del ataque fue un malware del tipo *Botnet*. Luego de infectar los equipos y propagarse en la red, ordenó el ataque. Las pérdidas ascendieron a cientos de millones de dólares.

Los algoritmos livianos *Block Ciphers* [12-14] o *Stream Ciphers* [14-18] pueden contribuir a mitigar los problemas de confidencialidad. Asimismo pueden utilizarse también algoritmos para la *Gestión de Claves, Firma Digital* y funciones *Hash* [19-21]. Y por supuesto, esto sin por ello reducir la fortaleza criptográfica que de ellos se espera.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La comunidad científica ya cuenta con muchos algoritmos criptográficos livianos. Algunos de ellos fueron creados por investigadores de universidades. Otros fueron desarrollados por empresas que requerían que sus equipos adquieran mecanismos de seguridad. Una parte del proyecto se destina al relevamiento, estudio y análisis de estos últimos algoritmos, sobre todo los que fueron ideados para dispositivos IIoT.

Además, se destina esfuerzos al análisis de los protocolos de comunicaciones en IoT que utilicen criptografía y la búsqueda de sus puntos débiles o vulnerables.

Finalmente, el proyecto sigue de cerca la última etapa del concurso del *NIST* [22] que pretende establecer el estándar criptográfico de cifrado autenticado de Criptografía Liviana para dispositivos IoT. Se espera que en breve se conozca al algoritmo finalista, que mediante el proceso de estandarización correspondiente, ofrezca sus servicios.

Por supuesto, apenas eso ocurra, el equipo procederá a su estudio, análisis y difusión del mismo, tal como se ha informado anteriormente.

3. RESULTADOS OBTENIDOS/ ESPERADOS

Estudios previos llevados adelante por el mismo equipo de investigadores del presente proyecto, ha llevado adelante un profundo relevamiento de muchos de los algoritmos empleados en IoT e IIoT, para detectar que la mayoría de ellos poseen vulnerabilidades que han permitido debilitar o romper la seguridad que ofrecían [23-24].

Por lo tanto, es recomendable el seguimiento de los algoritmos criptográficos, estudio y análisis de los mismos, dado que en cualquier momento pudiera detectarse un ataque. Y lo que se consideraba altamente seguro, deja de serlo.

El seguimiento y difusión de tales resultados, debe considerarse en gran medida, una tarea permanente.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de investigadores pertenece al cuerpo docente de *Tecnologías Aplicadas* en la *Facultad de Ingeniería*, el área de la *Seguridad Informática*, de la *Universidad del Salvador*. A ellos se suma una alumna para hacer su pasantía. Ella se encuentra promediando la carrera de *Ingeniería en Informática*, en nuestra Facultad.

Se espera que en el presente año el equipo pueda crecer con la incorporación de más docentes investigadores y alumnos. Incluso poder tener alumnos que quieran llevar adelante su Proyecto de Promoción y Síntesis, en el área de la Criptología.

La incorporación de docentes y alumnos redundará en un aumento del activo académico e investigativo para la unidad académica, como así también sembrando las bases para la investigación del futuro, a través de la participación de alumnos, para beneficio de nuestra universidad.

5. REFERENCIAS

- [1] Shancang Li , Lida Xu, Securing the Internet of Things. Syngress Media,U.S. Rockland, MA, United States. 2017.
- [2] Autor no informado. “Buenos Aires, una ciudad con iluminación inteligente”. Portal de información y trámites de la Ciudad Autónoma de Buenos Aires. Mayo, 2015. <https://www.buenosaires.gob.ar/noticias/buenos-aires-una-ciudad-con-iluminación-inteligente>
- [3] Autor no informado. Portal CONSTRUAR, Periódico Digital de la Construcción, propiedad de Gómez Nieto Consultores Asociados <https://www.construar.com.ar/2016/01/las-autopistas-de-buenos-aires-estrenan-iluminacion-inteligente/>
- [4] Mármol, H. “Cuáles son y qué hacen las ciudades argentinas que quieren parecerse a Japón” Portal del Diario Clarín, Septiembre 2019. https://www.clarin.com/tecnologia/smart-cities-hacen-ciudades-argentinas-quieren-parecerse-japon_0_i0n7KiJ5K.html.
- [5] ISO/IEC 29192. Information Technology - Security Techniques - Lightweight Cryptography. 2012.
- [6] Panasenko, S.; Smagin, S. “Lightweight Cryptography: Underlying Principles and Approaches”. International Journal of Computer

- Theory and Engineering, Vol. 3, No. 4, August 2011.
- [7] Manyika, J.; Chui, M.; Bughin, J.; Dobbs, R.; Bisson, P.; Marrs, A. "Disruptive technologies: Advances that will transform life, business, and the global economy". McKinsey Global Institute. 2013.
- [8] Evans, D. "Internet of Things La próxima evolución de Internet lo está cambiando todo". Cisco IBSG. 2012.
- [9] Román R., Nájera P., López J. "Los Desafíos De Seguridad En La Internet De Los Objetos" University.
- [10] Fei Hu, Security and Privacy in Internet of Things (IoTs): Models, Algorithms, and Implementations. Taylor & Francis Inc. Portland, United States. 2016.
- [11] Masanobu Katagi; Shiho Moriai, Lightweight Cryptography for the Internet of Things; Sony Corporation; 2016.
- [12] Satoh, A.; Morioka, S. "Hardware-Focused Performance Comparison for the Standard Block Ciphers AES, Camellia, and Triple-DES". Conference: Information Security, 6th International Conference, ISC 2003, Bristol, UK, October 1-3, 2003, Proceedings.
- [13] Beaulieu, R.; Shors, R.; Smith, J.; Treatman-Clark, S.; Weeks, B.; Wingers, L. "The SIMON and SPECK Families of Lightweight Block Ciphers." Cryptology EPrint Archive. International Association for Cryptologic Research, 19 June 2013.
- [14] Dworkin, M. "NIST SP 800-38B, Recommendation for Block Cipher Modes of Operation: The CMAC Mode for Authentication." NIST Computer Security Resource Center. National Institute of Standards and Technology, Spring (2005).
- [15] Daniel J. Bernstein. "The Salsa20 family of stream ciphers" . URL:<http://cr.yp.to/papers.html#salsafamily>. (2007).
- [16] Babbage, S.; Dodd, M. "The MICKEY stream ciphers". In New Stream Cipher Designs. Pp. 191-209. Springer Berlin Heidelberg. (2008).
- [17] Hell, M.; Johansson, T.; Meier, W. "Grain: a stream cipher for constrained environments". International Journal of Wireless and Mobile Computing, 2, pp. 86-93 (2007).
- [18] De Canniere, C.; Preneel, B. "Trivium. New Stream Cipher Designs (pp. 244-266). Springer Berlin Heidelberg. (2008).
- [19] Kavun, E. B., & Yalcin, T. "On the suitability of SHA-3 finalists for lightweight applications". The Third SHA-3 Candidate Conference. (2012).
- [20] Hirose, S., Ideguchi, K., Kuwakado, H., Owada, T., Preneel, B., & Yoshida, H. "A lightweight 256-bit hash function for hardware and low-end devices: Lesamnta-LW". International Conference on Information Security and Cryptology. Pp. 151-168. Springer Berlin Heidelberg (2010).
- [21] Guo, J.; Peyrin, T.; Poschmann, A. "The PHOTON family of lightweight hash functions". Advances in Cryptology-CRYPTO 2011 (pp. 222-239). Springer Berlin Heidelberg (2011).
- [22] <https://csrc.nist.gov/projects/lightweight-cryptography>
- [23] Eterovic, J.; Cipriano, M.; García, E.; Torres, L. Criptografía Ligera en Internet de las Cosas para la Industria. Congreso Argentino de Ciencias de la Computación. CACIC 2019. Libro de Actas. Pág. 1228-1240. UniRío. ISBN 978-987-688-377-1. 2019.
- [24] Eterovic, J. Cipriano, M. García, E. Torres, L. Lightweight Cryptography in IIoT The Internet of Things in the Industrial field. Computer Science Cacic 2019. Revised Selected Papers. Springer. ISBN 978-3-030-48324-1.

TD – Tesis Doctorales

INFORMÁTICA FORENSE: MÉTODOS, HERRAMIENTAS Y TÉCNICAS

Susana Herrera, Liliana Figueroa, Cecilia Lara, Graciela Viaña, Analía Méndez, Lilia Palomo, Luis Pianazzola

Instituto de Investigación en Informática y Sistemas de Información, Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero
sherrera@unse.edu.ar; lmvfigueroa@yahoo.com.ar; laraceciliacristina@gmail.com;
gv857@hotmail.com; anmendez725@yahoo.com; lilia.palomo@gmail.com,
luispianazzola@gmail.com

RESUMEN

Resulta relevante destacar los desafíos que genera el avance tecnológico para incorporar la utilización de la evidencia digital al sistema procesal penal, como prueba fundamental en la investigación de cualquier delito. En la provincia de Santiago del Estero y de manera gradual desde el año 2016 se ha implementado el Sistema Penal Acusatorio, siendo entonces necesario disponer de métodos, herramientas y técnicas que permitan una gestión eficiente del ámbito de trabajo en donde se realizan las actividades de informática forense en los distintos organismos e instituciones involucrados en este sistema.

Entonces, es necesario contar con métodos científicos que permitan recolectar, analizar y validar pruebas digitales que sean legalmente aceptables y que ayuden a resolver la investigación penal.

En esta propuesta se pretende llevar a cabo una investigación aplicada para proponer métodos, técnicas y herramientas forenses tendientes a garantizar una gestión eficiente de los laboratorios de informática forense que forman parte de las distintas instituciones del Sistema Penal Acusatorio en la Provincia de Santiago del Estero.

Palabras clave:

Informática forense, evidencias digitales, métodos-técnicas-herramientas forenses, laboratorios de informática forense.

CONTEXTO

En este artículo se presenta una propuesta de investigación que constituye una

continuación de los proyectos “Computación Móvil: desarrollo de aplicaciones y análisis forense” y Métodos y Herramientas para el análisis forense”, financiados en 2017-2018 y 2019-2021 por el Consejo de Ciencia y Técnica de la Universidad Nacional de Santiago del Estero [4].

En estos proyectos se lograron, hasta la fecha resultados referidos al: análisis de la obtención legal de la evidencia digital en los códigos procesales de nuestro país, análisis de antecedentes jurisprudenciales sobre tratamiento de evidencia digital en dispositivos móviles, investigación y análisis de protocolos vigentes en otras jurisdicciones, propuesta de protocolo para la obtención de evidencia digital móviles en concordancia con las normas ISO/IEC 27037:2012 en el ámbito del Ministerio Público Fiscal de Santiago del Estero, y estudio de repositorios que permitan la construcción de un modelo de datos para el almacenamiento y la gestión de evidencias digitales extraídas de dispositivos móviles y evaluación sistémica del protocolo de actuación propuesto.

La implementación del sistema procesal penal acusatorio tiene que ver con un cambio de paradigma, reasignando los roles de los actores de la Justicia en el tratamiento de las causas. Con este sistema cambia la forma de llevar a cabo el proceso y recae en los fiscales la tarea de investigar.

En este sistema lo que se busca es ayudar a que todos los actores de la justicia tengan pleno conocimiento de lo que pasa en el proceso. [1] En particular, en Santiago del Estero rige el código de procedimiento penal Ley N° 6941 [9], donde la Investigación

Penal Preparatoria (IPP), es la etapa del proceso penal que tiene por objeto determinar la existencia de delitos y la individualización de los eventuales autores. En este contexto, el modo de vincular un hecho criminoso con su autor, es a través de pruebas. Es sabido que la mayoría de las IPP cuentan con trabajos de investigadores y peritos; pero también es sabido que los cambios que se produjeron en estos últimos años hace necesario encontrar las huellas en el ámbito digital para encontrar las evidencias digitales que ayuden a resolver crímenes.

Atendiendo a esta realidad, los actores del sistema procesal penal acusatorio de la provincia de Santiago del Estero han tomado la iniciativa de implementar áreas específicas relacionadas a la informática forense, con el objetivo de dar respuesta a distintos ilícitos que tengan asociados dispositivos tecnológicos. En este sentido, se puede observar distintos niveles de avance en la gestión, mientras que algunos tienen incorporado en sus funciones un laboratorio de informática forense donde se han definido algunos procesos en relaciones a las tareas periciales, otros en cambio están en un proceso inicial de gestión y solamente brindan apoyo técnico a través de sus oficinas de informática.

En este ámbito se integrarán el personal especializado, la infraestructura física y tecnológica, las herramientas de hardware y software adecuadas para el análisis de datos, con el fin de recolectar evidencias que cumplan los principios de admisibilidad y tengan validez en el proceso judicial [3].

Entonces, la creación, operación y organización de laboratorios judiciales dedicados a la realización de pericias informáticas es una temática en donde aún se demandan varios aspectos para lograr una gestión eficiente de los mismos. En este contexto se pueden identificar diferentes problemas tanto administrativos como los que se enfrentan los operadores judiciales y peritos informáticos:

- Algunos procesos vinculados a la obtención de la evidencia digital tienen una

fuerte descripción desde una visión técnica, propiciando en algunos casos riesgos legales que garanticen su validez.

- Algunos organismos han ido incorporando estos espacios funcionales a partir de la incorporación de la infraestructura física, tecnológica y de las herramientas de software, sin una definición inicial de procesos que sostengan las actividades y tareas que se desarrollan.

- Se especifican procedimientos sin definir un acto administrativo resolutorio.

- Algunos de los organismos no cuentan con una estructura funcional y una descripción de los distintos puestos de trabajo, responsabilidades, perfiles para ocupar los puestos de trabajos.

- El costo elevado y la disponibilidad presupuestaria limita la adquisición de software especializado para la realización de las tareas de los peritos.

- La sobrecarga de las tareas diarias en relación a las pericias limita a los profesionales de disponer tiempo suficiente para diseñar casos de estudio de situaciones hipotéticas que propicien el uso eficiente de las herramientas en distintas situaciones de la investigación criminal.

- Las experiencias y las capacitaciones que se adquieren durante la realización de las pericias no se documentan para fomentar el autoaprendizaje organizacional.

- En los últimos años se ha incrementado el volumen de pericias sobre dispositivos tecnológicos y se plantean los siguientes inconvenientes:

- ✓ El proceso de obtención y análisis de las evidencias digitales es una tarea compleja que demanda un tiempo que en algunos casos excede los plazos procesales.

- ✓ No hay un software de informática forense que soporte la extracción de datos de todos los dispositivos móviles existentes en el mundo.

- ✓ Los requerimientos para los peritos forenses no están claros y definidos desde el inicio.

- ✓ Tomando como referencia el “Sistema de clasificación de herramientas forenses de

dispositivos móviles” desarrollado por Sam Brothers [2], que presenta cinco niveles sobre la extracción y el análisis requerido en cada dispositivo móvil teniendo en cuenta la solicitud y los detalles de la investigación. A partir de estos problemas, en esta propuesta se pretende investigar sobre estrategias que permitan gestionar de manera eficiente los distintos recursos involucrados en el entorno de los laboratorios de informática forenses, de manera tal que promueva el trabajo eficiente de la labor de los peritos ofreciendo un respaldo jurídico a su tarea.

1. INTRODUCCIÓN

La Informática forense es una rama de las ciencias forenses, que involucra la aplicación de la metodología y la ciencia para identificar, preservar, recuperar, extraer, documentar e interpretar [15] evidencias procedentes de fuentes digitales con el fin de facilitar la reconstrucción de los hechos encontrados en la escena del crimen [13], para luego usar dichas evidencias como elemento material probatorio en un proceso judicial [5] [4]; de esta manera constituye una disciplina auxiliar a la justicia, que consiste en la aplicación de técnicas que permiten adquirir, validar, analizar y presentar datos que han sido procesados electrónicamente y guardados en un medio computacional. [6] Por otra parte, el FBI, considera que la Informática forense es la ciencia que se encarga de adquirir, preservar, analizar y presentar los datos que han sido procesados electrónicamente y almacenados en medios electrónicos aplicando técnicas científicas y analíticas, utilizando hardware y software especializado para realizar la tarea. [10]

En este contexto los elementos informáticos que se disponen en los celulares, las computadoras o los dispositivos de almacenamiento como imágenes, mensajes de texto, conversaciones de WhatsApp, videos, ubicación, etc. constituyen las evidencias necesarias para resolver diferentes tipos de casos judiciales. Todos estos elementos son trabajados en los Laboratorios de informática forense, en donde se realizan las extracciones forenses

apoyadas con aparatología especializada y programas forenses; luego la información obtenida es procesada y analizada para conseguir la evidencia digital necesaria para las investigaciones. Por otra parte, en estos laboratorios no solo hay que recuperar la información, sino también realizar procedimientos de la forma adecuada para poder utilizar esta información como evidencia; entonces hay que garantizar la confiabilidad de la prueba y la cadena de custodia de la información, cuando se realiza una operatoria de estas características.

Una propuesta de los procedimientos para el establecimiento y gestión de Laboratorios informática forense son las enunciadas por [8], estas directrices proporcionan técnicas para el manejo y procesamiento de pruebas digitales; tienen como objetivo proporcionar un marco universal para establecer y administrar un Laboratorio de informática forense. Además, también se debe tener en cuenta el cumplimiento de principios básicos definidos por [7] para recoger y manipular pruebas digitales, estos principios sirven como marco de referencia y apoyan los procedimientos de actuación que se desarrollen en dichos laboratorios.

Otro aspecto importante que debe plantearse en los Laboratorios de informática forense a la hora de hacer el análisis de las evidencias, es la disponibilidad de herramientas forenses, entre las alternativas existen diferentes soluciones tanto pagas como gratuitas para este fin.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Considerando la amplitud de los aspectos relacionados con la informática forense, en el presente proyecto de investigación se diferencian las siguientes líneas de investigación:

- *Estudio de herramientas de informática forense,*
- *Desarrollo de aplicaciones de apoyo a la gestión de los laboratorios de informática forense.*

- *Técnicas y métodos para la gestión de laboratorios de informática forense.*

3. OBJETIVOS

El objetivo general de la investigación propuesta es:

Contribuir al mejoramiento de la gestión de las evidencias digitales en el ámbito del sistema judicial de Santiago del Estero.

Los objetivos específicos por cada línea de investigación que permitirán alcanzar el objetivo general son:

- *Estudio de herramientas de informática forense.*

a Estudiar comparativamente las arquitecturas de los dispositivos móviles.

b Evaluar herramientas de informática forense correspondientes a los niveles (1, 2 y 3) de la *Pirámide Móvil Forense*.

c Estudiar sistemáticamente herramientas correspondientes a los niveles superiores (4 y 5) de la *Pirámide Móvil Forense*.

- *Desarrollo de aplicaciones de apoyo a la gestión de los Laboratorios de informática forense.*

d Desarrollar una aplicación que de soporte a la formación inicial de peritos informáticos.

e Desarrollar nuevos módulos de la aplicación de gestión de experiencias de los peritos informáticos.

f Desarrollar un repositorio institucional para alojar evidencias digitales, permitiendo la búsqueda, visualización y recuperación de las mismas.

- *Técnicas y métodos para la gestión de laboratorios de informática forense.*

g Proponer modelos de gestión a las instituciones del sistema judicial de Santiago del Estero sobre la definición de normas, procesos y procedimientos.

Se trata de una investigación aplicada, desarrollada desde un enfoque cuantitativo, en el campo de la Informática Forense.

La investigación es aplicada porque los resultados que se obtengan se aplicarán inmediatamente la oficina de Informática Forense del Gabinete de Ciencias Forenses del Ministerio Público Fiscal, en el Gabinete

de Informática Forense del Poder Judicial de la Provincia de Santiago del Estero, en la Policía de la provincia de Santiago del Estero, en la Oficina de Informática del Ministerio Público de la Defensa.

4. FORMACIÓN DE RECURSOS HUMANOS

En cuanto a los RRHH, se espera consolidar y agrandar el equipo de investigadores de UNSE, iniciado en el año 2017. Se trabajará en forma conjunta y colaborativa con la oficina de Informática Forense del Gabinete de Ciencias Forenses del Ministerio Público Fiscal, con el Gabinete de Informática Forense del Poder Judicial de la Provincia de Santiago del Estero, la Oficina de Informática del Ministerio Público de la Defensa y en la Policía de la provincia de Santiago del Estero.

La concreción de este proyecto contribuirá a la formación y capacitación de los investigadores, estudiantes de grado involucrados. Incentivará a los alumnos a iniciarse en las actividades de investigación y favorecerá la realización de sus trabajos finales de grado en las líneas de investigación de este proyecto

Los investigadores constituyen un equipo interdisciplinario conformado docentes de la UNSE e investigadores externos, de profesión en Informática, Electromecánica y Derecho.

5. REFERENCIAS

- [1] Argentina.gob.ar. (2021, 6 de septiembre). Sistema acusatorio: nuevos roles y mayor eficacia. <https://www.argentina.gob.ar/noticias/sistema-acusatorio-nuevos-roles-y-mayor-eficacia>
- [2] Ayers, R; Brothers, S; Jansen, W. (2014). Guidelines on Mobile Device Forensics. NIST Special Publication 800-101. Revisión 1. <https://nvlpubs.nist.gov/nistpubs/specialpublications/nist.sp.800-101r1.pdf>
- [3] Calderón Valdiviezo RG, Guzmán Reyes GS, Salinas González JM, Aranda A. Diseño y Plan de Implementación de un Laboratorio

de Ciencias Forenses Digitales. (2011). file:///C:/Users/hp/Downloads/paper_laboratorio_forense_digital.pdf

[4] Castillo, C., Romero, A., Cano, J.: Análisis Forense Orientado a Incidentes en Teléfonos Celulares GSM: Una Guía Metodológica. Conf. XXXIV Conferencia Latinoamericana de Informática, Centro Latinoamericano de Estudios en Informática (CLEI). (2008). <http://www.clei2008.org.ar/>

[5] Del Pino, S.: Introducción a la informática forense. Pontificia Universidad Católica del Ecuador. (2007). <http://www.alfaredi.org/sites/default/files/articles/files/Acurio.pdf>

[6] Di Iorio, Ana H. -La informática forense y el Proceso de recuperación de Información digital. Revista Democracia Digital e Governo Eletrônico (ISSN 2175-9391), n° 8, p. 326-339 - 2013.

[7] Digital Evidence: standards and principles”, Forensic Science Communications, FBI, apr. 2000. <https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/april2000/swgde.htm>

[8] Directrices Mundial de INTERPOL para los Laboratorios Forenses Digitales. INTERPOL L Complejo Mundial para la Innovación de 2019. [https://Downloads/INTERPOL_DFL_GlobalGuidelinesDigitalForensicsLaboratory%20\(2\).pdf](https://Downloads/INTERPOL_DFL_GlobalGuidelinesDigitalForensicsLaboratory%20(2).pdf)

[9] Ley provincial 6.941. Código Procesal Penal de la Provincia de Santiago del Estero (2009).

[10] Noblett M. G., Pollit, M. M. “FBI” FBI, October 2000. <https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/oct2000/computer.htm>

[11] Orta Martínez, R.: Informática Forense como Medio de Pruebas. <https://www.dragonjar.org/informatica-forense-como-medio-de-pruebas.xhtml>

[12] Poder Judicial de Neuquén: Pericias informáticas sobre telefonía celular. (2013). <http://200.70.33.130/images2/Biblioteca/ProtocoloPericiasTelefoniaCelular.pdf>

[13] Reith, M., Clint, C., Gunsch G.: An Examination of Digital Forensic Models. International Journal of Digital Evidence, Air Force Institute of Technology, Volume 1 Issue 3. (2002). www.utica.edu/academic/institutes/ecii/publications/articles/A04A40DC-A6F6-F2C1-98F94F16AF57232D.pdf.

[14] Unidad Fiscal Especializada en Ciberdelincuencia: Guía de obtención, preservación y tratamiento de evidencia digital. (2016). <https://www.fiscales.gob.ar/procuracion-general/wp-content/uploads/sites/9/2016/04/PGN-0756-2016-001.pdf>

[15] Zdziarski, J.: iPhone Forensics, Recovering Evidence, Personal Data & Corporate Assets. O’Reilly Media, Inc. (2008)

Hybrid Networking SDN y SD-WAN: Interoperabilidad de Arquitecturas de Redes Tradicionales y Redes definidas por Software en la era de la digitalización

Gustavo D. Salazar-Chacón

Tesis Doctoral / Secretaría de Posgrado / Facultad de Informática
Universidad Nacional de La Plata, UNLP

RESUMEN

Software-Defined Networking (o SDN por sus siglas en inglés), desde su creación y posterior adopción, ha prometido ser la solución a los problemas de gestión, configuración y desempeño de las infraestructuras de red a través del uso de técnicas como programabilidad, *Hardware Abierto* con capacidades programáticas, agilidad extrema, además del empleo de Interfaces Gráficas/APIs seguras que otorguen visibilidad total de la infraestructura.

La presente tesis doctoral da un vistazo a la evolución de las redes de datos hacia el paradigma SDN y sus diversas adopciones (*SD-Access*, *SD-Data Center* y *SD-WAN*) con el fin de comprobar su facilidad de implementación, para lo cual se aborda los fundamentos de dichas tecnologías, partiendo de lo que implica el desacoplamiento del Plano de Control del Plano de Datos en equipos de red, hasta el concepto de cambio cultural y tecnológico denominado *NetDevOps*, imprescindible para que el ecosistema ágil SDN funcione de forma adecuada, atravesando por el análisis de protocolos estandarizados de próxima generación que permitan la implementación de estos entornos en redes reales: LISP, VXLAN, OMP y *Segment-Routing*, desarrollando a la par pruebas de concepto (PoCs) en ambientes de emulación y con equipos físicos, cerrando de esa manera el proceso investigativo que da validez a la integración de SDN basadas en programabilidad con redes tradicionales, siendo esa justamente la mayor contribución entregada por esta tesis.

Palabras Clave: *SDN; SD-WAN; SD-Access; LISP; VXLAN; Segment-Routing; Viptela; Open Networking; NetDevOps; OpenFlow; Controladores SDN; REST-APIs.*

APORTES DE LA TESIS

Realizar pruebas de concepto sobre SDN, automatización y redes programables, así como de telemetría, *Open-Networking* y nuevos protocolos de transporte, planteando además modelos innovadores de interoperabilidad entre redes tradicionales y redes definidas por software son los aportes principales de esta investigación, pues permitieron llegar a las conclusiones finales luego de un análisis profundo de los resultados obtenidos en la fase de simulación/emulación y pruebas en equipos físicos, dando lugar a una red híbrida IP-SDN donde no necesariamente se tienen los equipos robustos o flujos de tráfico tan altos como en las redes de un Centro de Datos SDN (*SD-DC*).

Una vez verificada la factibilidad de los conceptos y protocolos planteados en la investigación mediante *software* emulador, la fase final del desarrollo de esta tesis doctoral comprueba el comportamiento de SDN en equipos reales, aterrizando los conceptos de programabilidad y SDN a la realidad.

INTRODUCCIÓN

Desde hace una década, las redes de datos empresariales han experimentado una evolución vertiginosa dada por la adopción de modelos basados en la nube, contenedores y microservicios, primordialmente por que entregan mayor flexibilidad, efectividad y reducción de costos en cuanto a Tecnologías de la Información se refiere, sin embargo, los requerimientos de usuarios y empresas actuales

exigen cada vez una mejor respuesta de la infraestructura de telecomunicaciones. Esos requerimientos fueron más evidentes durante la época de pandemia COVID-19, llevándonos a una “Era de Datos y Sabiduría”, virtualidad y Transformación Digital, ya que la mayor parte de las actividades humanas y económicas dependen de una adecuada y segura transferencia de información, es así como los datos se convirtieron en el bien intangible más importante del siglo XXI.

A medida que más dispositivos móviles 4G/5G-WiFi6E ingresan a la red, así como aplicaciones y servicios novedosos entran en funcionamiento, una mayor carga de envío y procesamiento se añade a la red, generando dificultades tanto en la administración y monitoreo, como en la configuración y restablecimiento en caso de fallas.

La infraestructura de redes de comunicaciones que da cabida a estos cambios tecnológicos también debe adaptarse a la hiperconectividad y Centros de Datos de nueva generación, por ello, muchos protocolos y tecnologías de transporte surgieron a la par de las necesidades del usuario final.

La Figura 1 muestra esa evolución, marcando a MPLS (*Multiprotocol Label Switching*) como punto de inflexión en la evolución de las redes de transporte de datos WAN, mientras la Figura 2 muestra una topología de SD-WAN como presente y futuro de las Redes de Transporte.

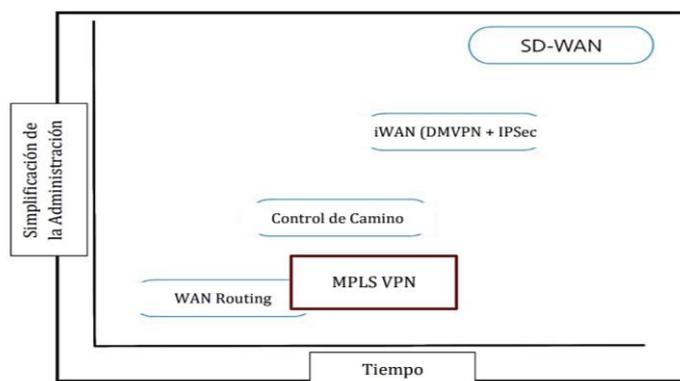


Figura 1 Evolución de las Redes WAN

Fuente: Autor

Las redes modernas deben tener ciertas características que complementen el criterio tradicional de escalabilidad, tolerancia a fallas, QoS y seguridad, pues los requerimientos de los usuarios y de las empresas así lo demandan.

Además de esas características básicas, deben tener:

- Resolución automática de fallas y errores.
- Defensa automática ante ataques.
- Optimización automática.
- Análisis automático de datos para una toma de decisiones adecuada.
- Visibilidad completa de la red mediante *Insights*.

El concepto de automático significa dotar de inteligencia y facilidad de manejo a las redes para que puedan adaptarse al entorno y de esa manera cubrir los requerimientos de IoT, *BigData*, BYOX, Conectividad 5G/6G, teletrabajo, *streaming* de video masivo, realidad virtual, entre otros.

Debido a que las redes requieren de un mejor entendimiento del contexto para generar visibilidad total y demás funciones importantes en las redes actuales, la importancia de los datos toma un nivel superlativo.

La digitalización da una importancia relevante a los datos, es más, todo elemento que forma parte de una red podrá generar algún tipo de información.

Según Juniper Networks, el tráfico de datos se incrementará en 9.6 veces para el año 2025, conectando una cantidad de más de 30 billones de dispositivos para el 2020-2022 y cerca de 80 billones para el 2025 (Wexler, 2016). Más que seguro, esa expectativa se quedará corta debido a los requerimientos de conectividad generados por la pandemia COVID-19.

Por otro lado, la transformación digital va más allá de la digitalización, pues no solo es una nueva tendencia tecnológica, es en realidad una estrategia empresarial y cambio cultural muy rentable que involucra a la innovación de los mercados y giros de negocio, pensando en el beneficio de los usuarios finales y de la empresa.

Gracias a la digitalización, los datos pueden ser fácilmente accesibles a través de diversas plataformas, equipos e interfaces. Ahora es momento de dar valor e importancia a cada uno de esos datos generados, provocando así una gran rentabilidad y nuevas plazas de trabajo.

FUNDAMENTOS DE LAS REDES DEFINIDAS POR SOFTWARE (SDN)

Hoy en día se considera a SDN un tópico tecnológico de moda, ya que prácticamente todas las implementaciones de redes de estos últimos años abarcan alguna variante del concepto *Software-Defined*.

Lo más interesante de esta evolución en las redes es que si bien ya tiene algunos años tratando de darse un lugar en las infraestructuras y en muchos casos, logrando posicionarse de una manera muy adecuada en el mercado del *networking* empresarial, no existe en sí una definición única para SDN pues aún depende del fabricante que lo despliegue, sin embargo, en todas las definiciones está la búsqueda de mejorar patrones de rendimiento, seguridad, enfocado a la nube y visibilidad usando el *Internet-Fabric* como medio de transporte, reinventando de esa manera las redes de datos y comunicaciones. SDN se encuentra en una etapa de estandarización, proceso importante para su desarrollo y así cubrir los requerimientos en la era de la digitalización.

Contar con una red que sea ágil, programable, rentable, segura e inteligente es un requerimiento y entender a SDN, siendo este

cambio tecnológico el puntal de esta migración, toma una gran relevancia. Ya desde inicios del 2017, se realizaron predicciones indicando el impacto económico en las industrias que no realicen transformaciones digitales en sus negocios. Para el 2020, los negocios e industrias que sean innovadoras tomaron \$USD1.2 Trillones de dólares de aquellas que no lo sean (McCormick, 2016). Con la pandemia esa cifra sería aún mayor.

Para Cisco Systems, SDN se implementa mediante el modelo de *Intent-Based Networking*, donde la escalabilidad, automatización a través de programabilidad, visibilidad centralizada y optimización de las redes tanto de infraestructura o *Enterprise*, DC y *Cloud* público/privado son elementos infaltables (Cisco Systems, 2020). Otros *vendors* muy relevantes en el desarrollo de SDN como VMWare, se han preparado durante los últimos años para adoptar esta nueva ola de innovación mediante virtualizaciones especiales de la red o NVs (*Network Virtualizations*) ofreciendo maneras de implementar SDN de forma flexible y sin utilizar demasiados recursos, tanto computacionales como económicos. VMWare adquirió Nicira en 2012, uno de los mayores impulsores de la creación de *vSwitches*, desarrollando VMWare NV en forma de contenedores denominándolo VMWare NSX, compañía dedicada a SD-DC (VMWare SDX-Central, 2018). Por su parte, Juniper (Juniper Networks, 2019), Citrix (Citrix, 2019), Huawei (Huawei, 2019), Cisco Systems (Cisco Meraki SD-WAN, 2019), entre otros fabricantes adoptaron también el concepto de *Underlay-Overlay Networks*, motivando así el despliegue de SD-WAN (Figura 2).

SDN también se relaciona con el uso de equipos abiertos u *Open Hardware*.

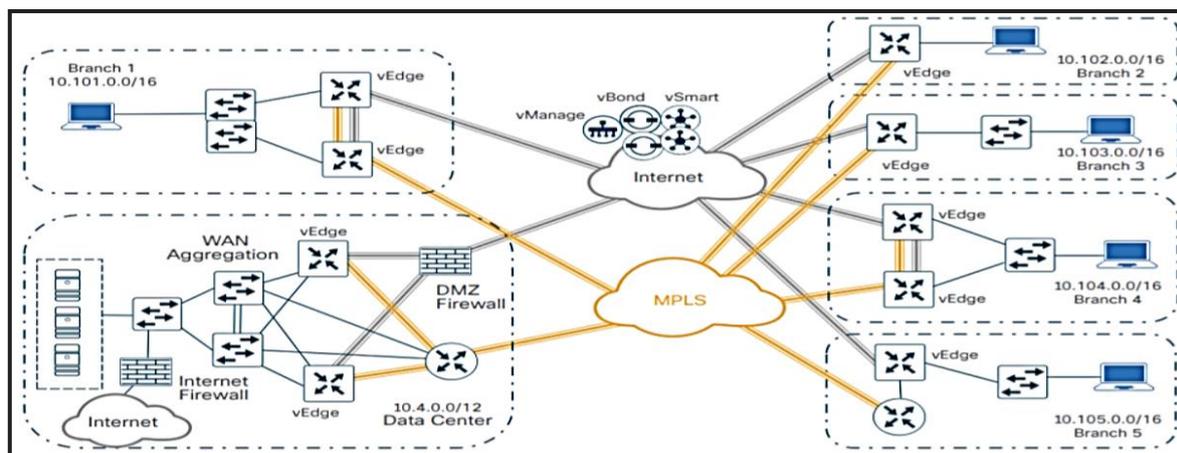


Figura 2 Infraestructura SD-WAN.
Recuperado de (Cisco Systems, 2019)

Debido a esta diversidad de conceptos, la ONF (*Open Networking Foundation*¹), consorcio sin fines de lucro creado para delinear el camino de los nuevos modelos de negocio e infraestructura, trabajando colaborativamente como una comunidad abierta para consolidar proyectos de desagregación de redes, *White-Box Networks*, así como para establecer estándares que revolucionarán la industria del *networking*, definió a SDN como la separación física del plano de control del plano de datos de un equipo, con el fin de centralizadamente controlar, monitorear y programar varios dispositivos a la vez a través de peticiones/respuestas desde una API (Figura 3).

Componentes de SDN

Los componentes básicos de una infraestructura de este tipo son:

➤ **Equipos de Red:** Son los dispositivos que conforman el plano de datos o plano de *forwarding* de la red. En este caso, estos dispositivos carecen de inteligencia y solo poseen rapidez de envío. Cabe mencionar que pueden ser equipos físicos o virtuales.

- **Controlador-SDN:** Equipo encargado de dotar de inteligencia a la red. Es el lugar donde se toman las decisiones de envío, se establecen políticas de enrutamiento, conmutación, seguridad, así como políticas empresariales. Conformar el Plano de Control centralizado de la red, es decir, su cerebro. Es posible que el controlador se comunique hacia las aplicaciones, así como hacia los equipos de red usando APIs especializados.
- **Northbound/Southbound Interfaces:** APIs que permiten la comunicación desde el controlador hacia las Aplicaciones y el Plano de datos respectivamente. Emplean protocolos de comunicaciones estandarizados o propietarios.
- **NOS (Network Operating System):** Sistema Operativo que emplean tanto los Controladores como los Equipos de Red. Pueden ser sistemas operativos abiertos o propietarios.

¹ ONF: <https://www.opennetworking.org/>

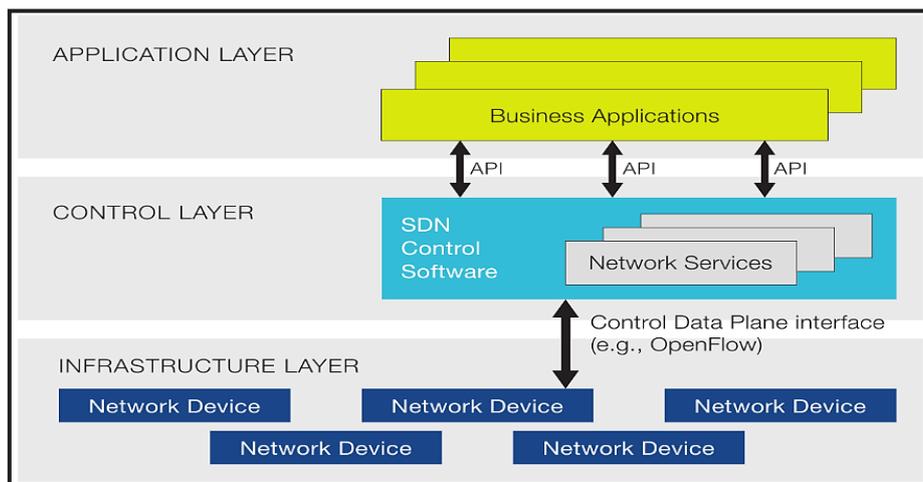


Figura 3 SDN según la ONF
Recuperado de (SDX Central, 2019)

PROGRAMABILIDAD DE REDES: *NetDevOps e IaaS*

SDN tiene el objetivo de crear infraestructuras mucho más ágiles, escalables, seguras y flexibles que las redes tradicionales. Uno de los mecanismos para llegar a ello es mediante la "Automatización" y así crear redes programables.

"Si el software va a comerse el mundo, entonces SDN se comerá al networking" fue dicho ya en el 2016 por (Rivenes, 2016). Las redes se han vuelto complejas y una forma de administrarlas adecuadamente es a través de SDN, es más, tal como lo hemos analizado, los usuarios demandan de cambios mucho más rápidos, velocidad que las redes tradicionales monolíticas no pueden satisfacer.

Según (Rivenes, 2016), los principios de SDN llevan las redes a lo que se acuñó como *Infrastructure-as-a-Code* (IaaS o IaC), lo cual permite que la automatización y orquestación sea una realidad, no obstante, IaaS también puede implementarse como una innovación en las redes tradicionales sin necesariamente llegar al punto de convertirla en SDN mediante mecanismos de automatización y programabilidad en redes y así exista una convivencia de redes modernas y de legado.

SDN y *DevOps* no son para nada excluyentes el uno del otro, en realidad, *DevOps* compagina perfectamente en estos entornos. *DevOps* se refiere a la orquestación de procesos complejos, interdependientes y asociados con el desarrollo de *software* y las Operaciones en Tecnologías de la Información con el fin de acelerar los procesos de producción y desarrollo de servicios. *DevOps* se trató algún tiempo atrás y nació desde el lado del manejo puro de IT (*Information Technology*).

La automatización de las redes ha sido un tópico requerido y anhelado por mucho tiempo, pero debido a su difícil implementación en redes tradicionales, no ha impactado sino en los últimos años. Junto con SDN y un ambiente tecnológico adecuado, de forma natural todo conducirá hacia una automatización en todos los niveles.

Partiendo de una investigación de la Revista *Network World* (Bednarz, 2016), se realizó una encuesta a 315 expertos en redes de empresas de tamaño mediano a grande, concluyendo que al rededor del 44% de los cambios que se generan en una infraestructura llevan a algún tipo de suspensión de servicios. La forma más habitual de encontrar y solucionar la red (mecanismo de *Tshoot*), según el 79% de los encuestados es usando CLI (*Command Line*

Interface) e inspeccionar la configuración manualmente, quizá usar algún tipo de herramienta de monitoreo SNMP y finalmente *ping* o realizar una traza (*traceroute*). El último dato obtenido de esta investigación fue que el tiempo promedio para dar respuesta a un problema, según el 60% de la población encuestada es de 1 a 5 horas.

NetDevOps es traer *DevOps* al *networking*; es así como un ingeniero en *DevOps* asegura la continua entrega del servicio IT, cambiando con ello el aspecto cultural-empresarial, buenas prácticas y herramientas que mejoran la efectividad. Bajo este modelo, tanto el equipo de desarrollo como el de operaciones no están separados, sino, se unen en un solo equipo de trabajo involucrándose en todas las etapas del ciclo de vida de una aplicación e infraestructura, por ello, las habilidades de los profesionales en este campo son variadas, algunas funcionando como eje transversal y llevando a la red a un concepto totalmente holístico.

En el modelo *DevOps*, existen tres prácticas habituales en el desarrollo, implementación y puesta en marcha de un *software*, una de las cuales se analizó en los párrafos anteriores:

- Integración Continua o CI por sus siglas en inglés de *Continuous Integration*: Práctica en la cual los desarrolladores guardan sus códigos en un repositorio central y colaborativo. En ese repositorio se pueden correr pruebas automáticas y gestión de versiones. Ejemplos muy importantes están en Git, Netbox, etc.
- Entrega Continua o CD por sus siglas en inglés de *Continuous Delivery*: Práctica donde los cambios en código son automáticamente enviados a un ambiente de pruebas después de la etapa de construcción (*Build Stage*).

El término CI-CD de igual manera se emplea para establecer este *pipeline*.

- IaaC o Infrastructure-as-Code: Práctica donde la infraestructura es aprovisionada y administrada usando código en oposición a usar CLI, empleando técnicas como CI y CD.

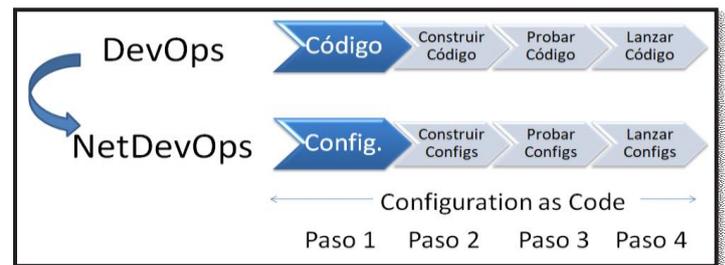


Figura 4 Cambios en una Red -NetDevOps Pipeline Basado en (Afaq, 2018)

APPLICATION PROGRAMMING INTERFACES: APIs

Los formatos de Datos son utilizados por distintos dispositivos inteligentes para transferir información relevante y realizar alguna actividad automatizada. La interfaz que permite ese intercambio se denomina API por las siglas en inglés de Interfaz de Aplicación Programable, es así como una definición rápida de API sería un software que permite a otras aplicaciones acceder a sus datos e interactuar con ellos.

Una API permite definir un conjunto de reglas que describen una aplicación y así interactuar con otra. Estas APIs, en el mundo hiperconectado actual, las encontramos por doquier, por ejemplo, AWS (*Amazon Web Services*), Facebook, Google, servicios de aerolíneas para buscar vuelos, o distintos entornos IoT utilizan algún tipo de API para interactuar con su entorno.

En esta era de la programabilidad, es necesario visualizar el entorno, el cual, bajo mi criterio está conformado por el Stack para programabilidad de redes (Figura 5).

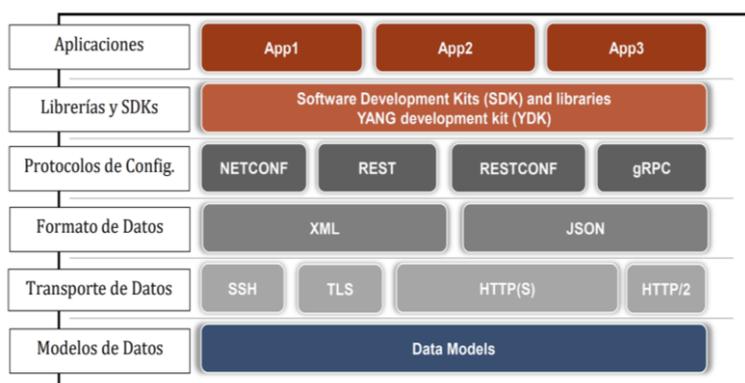


Figura 5 Stack para Programabilidad en las Redes
Recuperado de (Roman & Bryan, 2018)

RESTful API

REST es una API diseñada para arquitecturas de tipo servicios web, pero eso no implica que es únicamente utilizado con ese fin, en realidad, hace referencia al estilo de arquitectura web cliente-servidor, el cual posee algunas características y comportamiento en particular que puede ser ampliamente utilizado en *NetDevOps*.

En general, una REST API corre sobre el protocolo HTTP, definiendo un conjunto de peticiones-respuestas comunes llamadas métodos, dos conocidos son GET y POST.

Con el fin de implementar un RESTful API, es necesario tomar en consideración que, al implementarse mediante HTTP, se tienen cuatro (4) elementos básicos:

- URI (Identificador Uniforme de Recursos).
- Formato de datos admitido por la API (JSON, YAML, XML o cualquiera que sea un estándar de hipertexto válido).
- Métodos válidos para HTTP
- Control de API mediante hipertexto.

Partiendo de ese conocimiento, se puede usar mecanismos de solicitud-respuesta para orquestar, aprovisionar, configurar y monitorear de forma automática y programática a equipos y sistemas operativos de red (NOS).

Protocolos como NETCONF (*NET CONFIguration*) y RESTCONF, utilizan estas arquitecturas de intercambio de mensajes y dar mecanismos alternativos a la administración de una infraestructura de red moderna. En ese punto, el uso de un CLI (*Command Line Interface*) con protocolos de administración tradicional como SNMP es opcional, pues se podrá utilizar *scripts* en Python o usar cualquier herramienta que realice peticiones tipo RESTful.

Cuando se tiene en mente realizar programabilidad en una red, es imprescindible contar con un procedimiento claro de programación como el usado en la programación imperativa, el cual no es más que aquel enfocado a tener una secuencia de comandos ordenados con el fin de conseguir un objetivo, generando controles de flujo, verificación de estructura, telemetría, etc.

Uno de los objetivos primordiales de la automatización y programabilidad de redes es tener idempotencia en sus *scripts*, es decir, llegar a un estado deseado luego de la aplicación del código sin importar las condiciones iniciales en que se encuentre la infraestructura.

Herramientas de automatización y programabilidad como Ansible, Terraform, Puppet o Chef ofrecen capacidades muy poderosas en comparación con mecanismos de configuración y programación tradicionales como Bash o empleando nativamente algún lenguaje de programación como Python aún si se dispone de SDKs, pues estas herramientas reúnen funciones variadas en un entorno tipo API para simplificar y estandarizar su acceso, disminuyendo el hecho de incluir código extenso y tener *scripts* más simples y legibles, pero sin dejar a un lado la posibilidad de acceder al código puro para una revisión profunda.

Las herramientas antes mencionadas permiten además la inclusión de *plugins* totalmente configurables en *Python*, *Ruby*, *Go*, o cualquier otro lenguaje que sea aceptado.

Con el fin de lograr la comunicación entre el dispositivo central de configuración (o controlador de configuración) y la infraestructura de IT, es imprescindible un protocolo seguro que permita el envío y recepción de mensajes, inmediatamente, saltarán las siglas SSH (*Secure Shell*). **Paramiko**² es una implementación de SSH para Python usado para interconectar distintos *hosts*, sin embargo, resultó ser complicado para ciertos equipos en especial en procesos de autenticación, es así que surgió **Netmiko**³, una librería Python construida sobre Paramiko diseñada para solventar esas dificultades.

CONTROLADORES SDN Y *OpenSDN*

Durante el transcurso de la presente tesis, se ha dejado en claro que SDN es un nuevo paradigma para administrar, operar y enviar datos en una infraestructura de red, en la cual existe un controlador central encargado de brindar la inteligencia y toma de decisiones, liberando así de esa responsabilidad a los equipos, así como de la necesidad de intercambio de mensajes en el plano de datos, consiguiendo de esa manera mayor efectividad en la red.

El éxito de esta forma de envío de datos depende de que el Plano de Control esté separado y centralizado a través de un Controlador.

No obstante, SDN ya tiene algunos años de desarrollo, por ese motivo, los controladores han ido evolucionando en el tiempo. Se puede decir que existen dos conjuntos grandes:

- Controladores SDN para entornos NFV, DCs, WAN y Acceso modernos.

- Controladores SDN tradicionales para administrar el plano de Datos de una red.

Existe una lista de muchos controladores SDN que se adaptan a cualquiera de esos dos conjuntos, muchos incluso han servido como punto de partida de otros debido a la naturaleza mayoritaria de una comunidad *open-source*. En base al artículo de (Velrajan, 2019), la mayoría de los controladores SDN actuales funcionan bajo *OpenFlow* (*OpenSDN*).

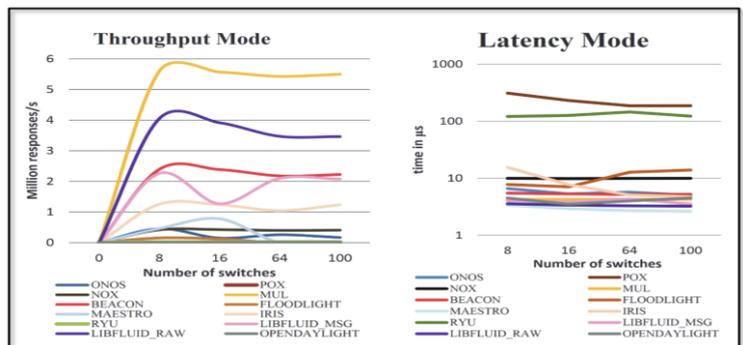


Figura 6 Comparación de Throughput y Latencia - Controladores SDN

Recuperado de (Salman, Elhadj, Kayssi, & Chehab, 2016)

PRUEBAS DE CONCEPTO (PoCs)

Se realizaron varias pruebas en emuladores de grandes prestaciones para corroborar la factibilidad de uso de estas nuevas tecnologías y sus APIs, finalizando con una prueba en equipos reales *OpenSDN* en ambientes cableados e inalámbricos.

Evolución SDN: NG-SDN y *Open-Hardware*

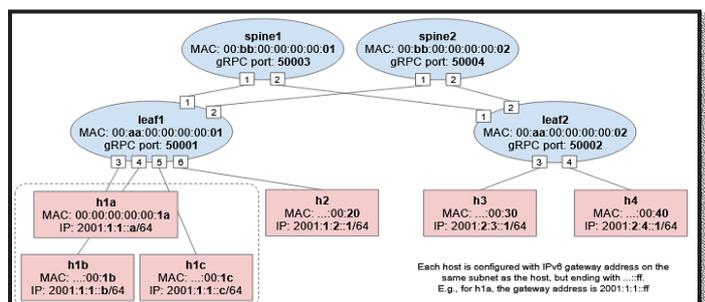


Figura 7 PoC - NG-SDN (uONOS y P4)

Fuente: Autor

² Paramiko: <http://www.paramiko.org/>

³ Netmiko: <https://pypi.org/project/netmiko/>

Evolución WAN - Segment-Routing

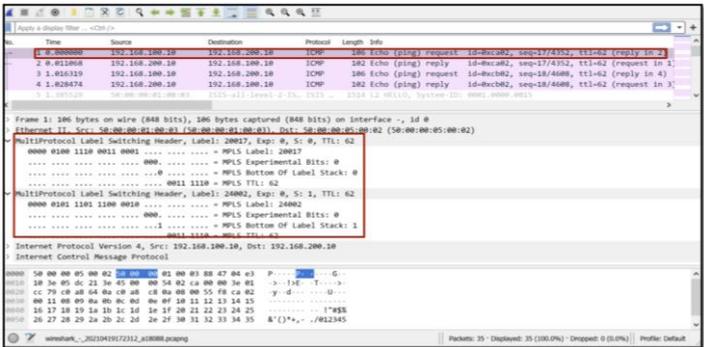
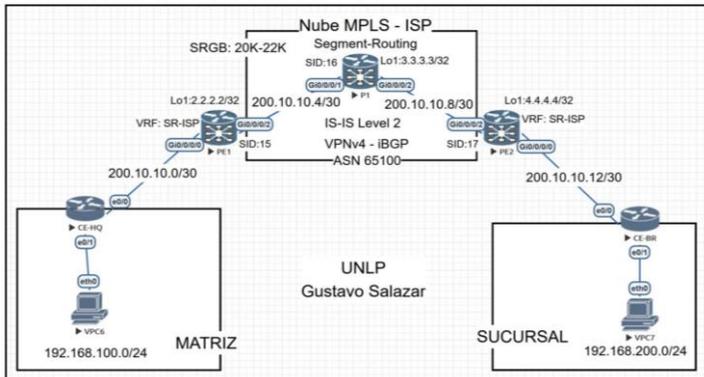


Figura 8 PoC – Segment-Routing
Fuente: Autor

OpenSDN Híbrido (ONOS)

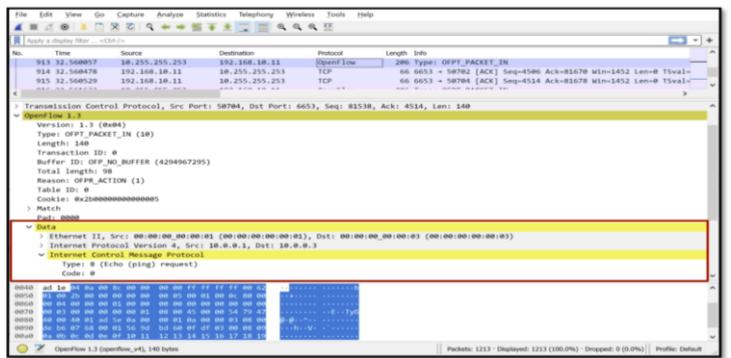
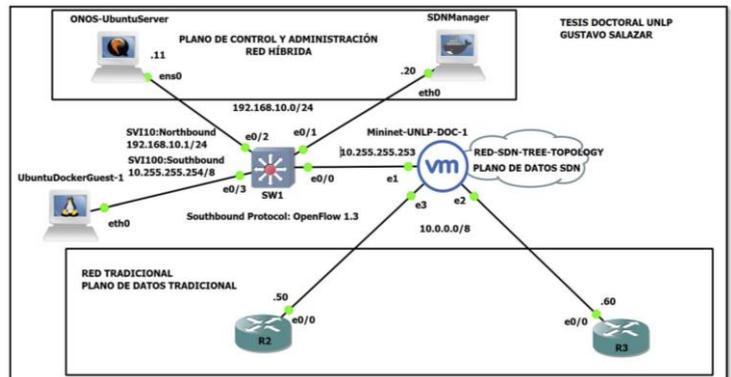


Figura 10 PoC – OpenSDN (ONOS)
Fuente: Autor

SD-Access: LISP y VXLAN con NetDevOps

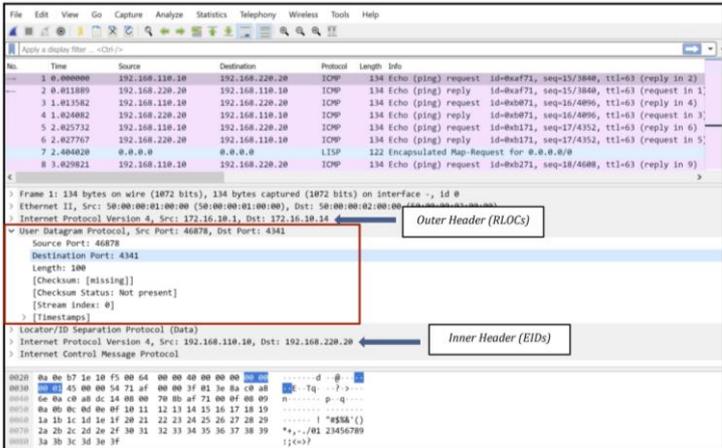
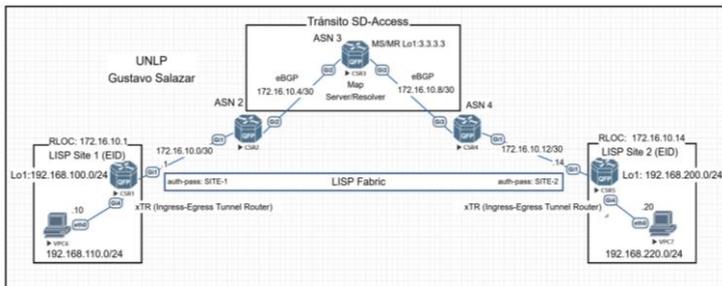


Figura 9 PoC – LISP
Fuente: Autor

SDWAN: Viptela - OMP

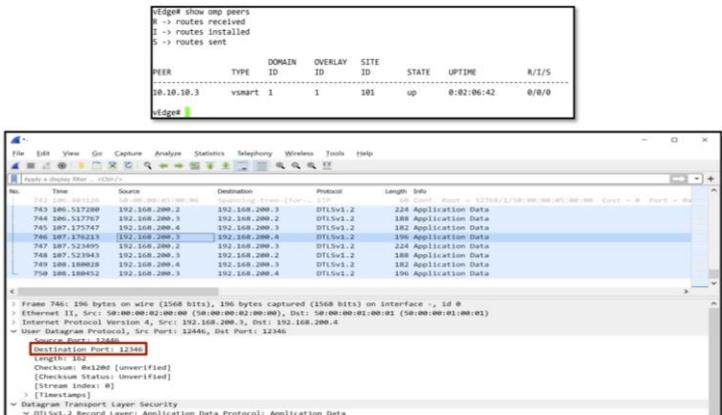
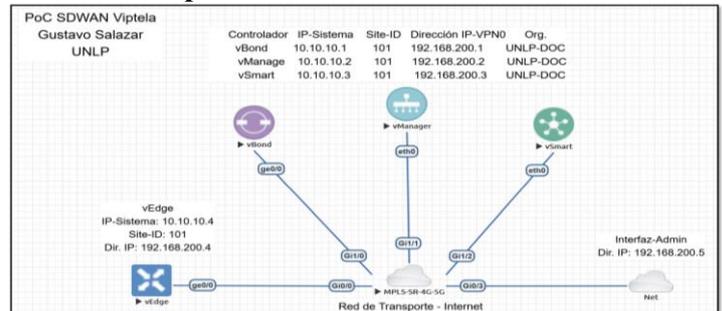


Figura 11 PoC – SDWAN (Viptela)
Fuente: Autor

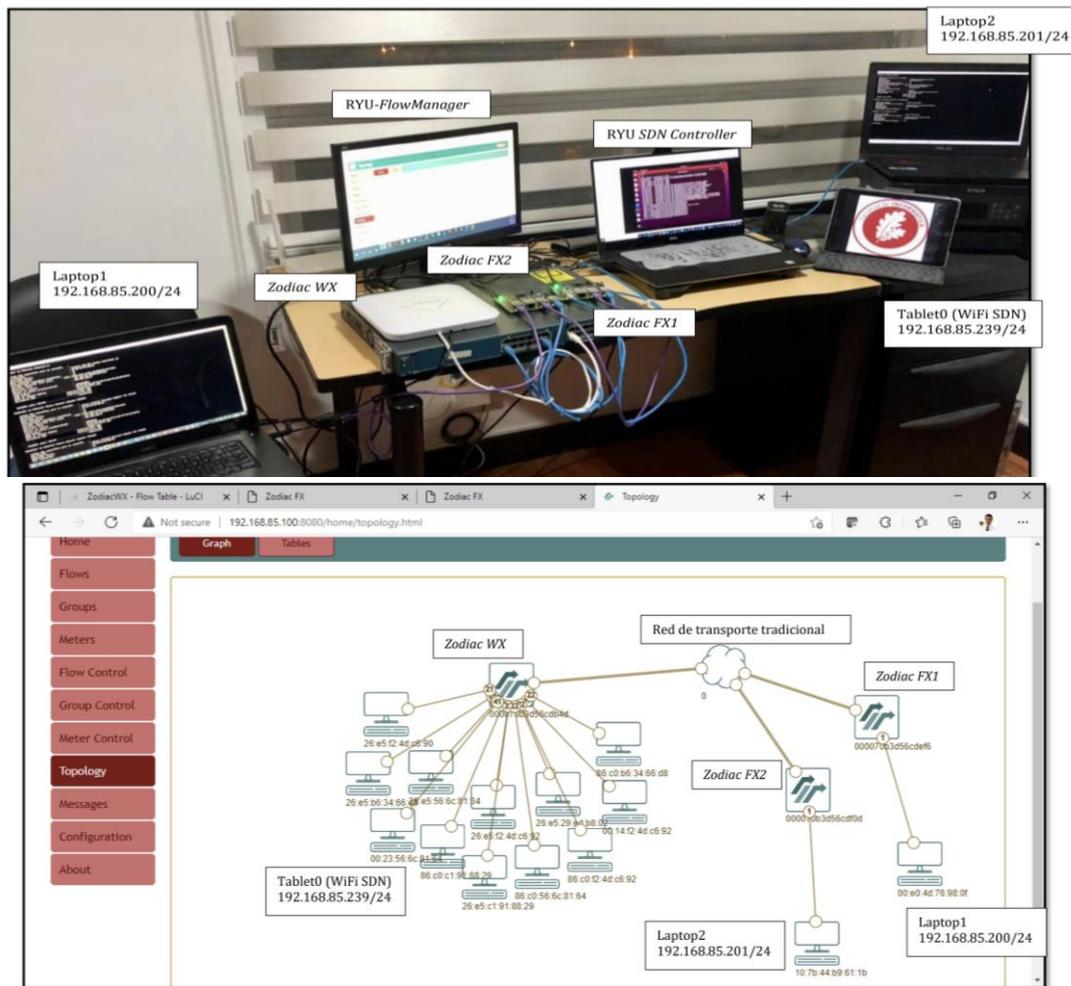


Figura 12 PoC – Prueba en Equipos Físicos OpenSDN Cableado e Inalámbrico (RYU)
Fuente: Autor

CONCLUSIONES

Luego de un proceso de investigación y posterior realización de pruebas de concepto bajo el principio de Desarrollo basado en Pruebas (*Test-Driven Development*), se identificaron los elementos clave para adaptar las redes tradicionales a la Era de la Programabilidad mediante un modelo de innovación propuesto:

- Redes seguras de extremo a extremo bajo el paradigma *Underlay-Overlay*, utilizando mecanismos de cifrado y autenticación robustos, con capacidad de conexión *multicloud* y entrega de servicios virtualizados. Este concepto crea la red híbrida IP-SDN, en el que las redes

tradicionales y SDN coexistirán de forma armónica e interoperable.

- Capacidad de orquestación, analítica y monitoreo (*insights*) a través de APIs en infraestructura abierta y *OpenNetworking* que reduzcan costos en CAPEX, OPEX y fomenten un pronto retorno de inversión (efectividad en las redes de datos).
- Uso de *frameworks* multiplataforma que faciliten la automatización de los procesos de redes, configuración y resolución de problemas (Programabilidad basada en Modelos), dando control de versiones de código y agilidad de implementación. El uso de patrones de diseño claros y estandarizados solucionan problemas

comunes siempre presentes en el desarrollo del *software* controlador de la red.

- Cambio cultural, empresarial, laboral, económico y tecnológico enfocado al uso de metodologías ágiles, *NetDevOps* y Transformación Digital coherente con la época de pandemia COVID-19.

Se sentaron las bases teóricas de SDN/*OpenSDN* según el modelo anteriormente planteado y bajo la premisa de *Separation of Concerns* (SoC) según el modelo MVC (*Model-View-Controller*), estudiando la estandarización, encapsulación y mecanismos de comunicación de *OpenFlow*, concluyendo que, a pesar de tener diferentes alcances y casos de uso, funcionan bien integrándolos en un ecosistema SDN de tipo CI/CD (Integración Continua / Implementación Continua).

Se analizó de igual manera a los principales controladores SDN al momento de la escritura de esta tesis con el fin de determinar a los mejores, llegando a la conclusión que ODL, RYU y ONOS tienen el mejor desempeño según los KPIs planteados (*throughput*, latencia, *jitter*, modularidad, soporte y capacidad de monitoreo) y pruebas de estrés (tasa de ráfaga). Cabe recalcar que los controladores desarrollados en C/Java como Mul, LibFluid y Maestro son más reactivos en entornos de estrés, pero menos adaptables a mejoras futuras, mientras los más modernos como *OpenDayLight*, su continua evolución y cambios en paqueterías, han complicado su integración con infraestructuras tradicionales.

Se estudió las distintas formas que puede tomar SDN en una infraestructura real corporativa: *SD-Access*, *SD-DC* y *SD-WAN*, manteniendo el precepto de desacoplamiento del Plano de Control del de Datos con el fin de mejorar la respuesta de redes de datos masivos y larga distancia a través del uso de protocolos de transporte/enrutamiento de próxima generación como LISP, VXLAN, *Segment-Routing* y

OMP. Su factibilidad de implementación e interoperabilidad con redes tradicionales fue comprobada con éxito en los PoCs establecidos.

La evolución hacia el ecosistema SDN será paulatino, aún nos encontramos en un momento de adaptación, es así que, el éxito de migrar a un entorno SD-WAN se medirá en su nivel de integración con redes tradicionales. OMP en SD-WAN Viptela es capaz de ser el puente entre las redes LAN y el *fabric* de SD-WAN, lo cual fue comprobado en los PoCs respectivos.

Finalmente, la factibilidad de implementación de *OpenSDN* (*OpenFlow*) en entornos reales mediante controlador RYU-*FlowManager* y Aruba VAN SDN, se corroboró con resultados prometedores, tanto en desempeño como facilidad de uso, lo que sin duda, deja un precedente, SDN no sólo sirve en entornos masivos de gran escala, también en entornos empresariales PyME-SMB donde los recursos de conectividad y procesamiento de datos son escasos, pero se requiere las mismas características básicas de una red bien diseñada empresarial: Escalabilidad, Seguridad, Calidad de Servicio y Resiliencia, sumando a ellas las ventajas de las Redes Definidas por *Software*: mejor desempeño, programabilidad y posibilidad de automatización para disminuir los tiempos de configuración y solución de problemas, además de reducir considerablemente el gasto operativo y de gestión de la red.

PUBLICACIONES RELACIONADAS CON LA TESIS Y BIBLIOGRAFÍA

- G. D. Salazar-Chacón and A. R. Reinoso García, "*Segment-Routing Analysis: Proof-of-Concept Emulation in IPv4 and IPv6 Service Provider Infrastructures*," 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 2021, pp. 1-7, doi:

- 10.1109/IEMTRONICS52119.2021.9422559.
- G. D. Salazar-Chacón and L. Marrone, "OpenSDN Southbound Traffic Characterization: Proof-of-Concept Virtualized SDN-Infrastructure," 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2020, pp. 0282-0287, doi: 10.1109/IEMCON51383.2020.9284938.
 - J. E. Vaca P. and G. D. Salazar-Chacón., "VXLAN-IPSec Dual-Overlay as a Security Technique in Virtualized Datacenter Environments" 2020 IEEE ANDESCON, 2020, pp. 1-6, doi: 10.1109/ANDESCON50619.2020.9272160.
 - G. Salazar-Chacón, E. Naranjo and L. Marrone. (2020). "Open networking programmability for VXLAN Data Centre infrastructures: Ansible and Cumulus Linux feasibility study". Revista Ibérica de Sistemas e Tecnologias de Informação, (E32), 469-482.
 - G. Salazar, "Ansible y SDN en acción: Los pilares de la Era de la Programabilidad", Cisco Community Support, 26 de noviembre, 2019 [online]. Disponible: <https://youtu.be/m0bSM8Xv10g>
 - G. D. Salazar Ch., C. Hervas, E. Estevez and L. Marrone, "High-Level IoT Governance Model Proposal for Digitized Ecosystems," 2019 International Conference on Information Systems and Software Technologies (ICI2ST), 2019, pp. 79-84, doi: 10.1109/ICI2ST.2019.00018.
 - G. D. Salazar Ch, C. Venegas and L. Marrone, "MQTT-Based Prototype Rover with Vision-As-A-Service (VAAS) in an IoT Dual-Stack Scenario," 2019 Sixth International Conference on eDemocracy & eGovernment (ICEDEG), 2019, pp. 344-349, doi: 10.1109/ICEDEG.2019.8734341.
 - G. D. Salazar Ch., E. F. Naranjo and L. Marrone, "SDN-Ready WAN networks: Segment Routing in MPLS-Based Environments," 2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2018, pp. 173-178, doi: 10.1109/UEMCON.2018.8796613. – Best Paper – Columbia University.
 - G. D. Salazar Ch., C. Venegas, M. Baca, I. Rodríguez and L. Marrone, "Open Middleware proposal for IoT focused on Industry 4.0," 2018 IEEE 2nd Colombian Conference on Robotics and Automation (CCRA), 2018, pp. 1-6, doi: 10.1109/CCRA.2018.8588117.
 - E. F. Naranjo and G. D. Salazar Ch, "Underlay and overlay networks: The approach to solve addressing and segmentation problems in the new networking era: VXLAN encapsulation with Cisco and open source networks," 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM), 2017, pp. 1-6, doi: 10.1109/ETCM.2017.8247505.
 - G. Salazar Chacón y G. Chafra Altamirano. (2015). "Empleo de Path-control Tools en una red empresarial moderna mediante Políticas de Enrutamiento". 3C Tecnología. Glosas De Innovación Aplicadas a La Pyme, 4(1), 1-18. Recuperado a partir de <http://ojs.3ciencias.com/index.php/3c-tecnologia/article/view/233>
 - G. Salazar, "DMVPN Fase1 y 2 en IPv4 Fundamentos y Configuración básica enfocado al CCIE RS", 2017, [online] Disponible: <https://community.cisco.com/t5/videos-routing-y-switching/dmvpn-fase1-y-2-en-ipv4-fundamentos-y-configuraci%C3%B3n-b%C3%A1sica/ba-p/3104173>
 - G. Salazar, "Fundamentos de IP Multicast Routing y sus Modos de Operación: Demo en Vivo", 2017, [online] Disponible: https://www.youtube.com/watch?v=xKD_Vppf8co

Diseño y Verificación de Sistemas de Tiempo Real Heterogéneos

Doctorado en Ciencias de la Computación

Departamento de Ciencias e Ingeniería de la Computación de la Universidad Nacional del Sur.

Autor: Francisco Ezequiel Páez (FI-UNPSJB) (fep@ing.unp.edu.ar)

Fecha de Defensa: 12/03/2021

Director: Javier D. Orozco (DIEC-UNS) (jadorozco@gmail.com)

Codirectores: José M. Urriza (FI-UNPSJB) (josemurriza@gmail.com) y Pablo Fillottrani (DCIC-UNS) (prf@cs.uns.edu.ar)

RESUMEN

La tesis aporta nuevos métodos de evaluación de planificabilidad y de administración del tiempo ocioso. Estos son utilizados en el diseño e implementación de Sistemas de Tiempo Real Heterogéneos, donde se requiere una planificación conjunta y eficientemente de tareas críticas y no-críticas. Los métodos propuestos se evalúan mediante simulaciones e implementaciones sobre placas de desarrollo y Sistemas Operativos de Tiempo Real, verificando así su factibilidad práctica.

Palabras clave: STR, Planificación, Slack Stealing.

CONTEXTO

Este documento es un resumen de la Tesis “Diseño y Verificación de Sistemas de Tiempo Real Heterogéneos”, presentada para la obtención del título de Doctor en Ciencias de la Computación, en el Departamento de Ciencias e Ingeniería de la Computación de la

Universidad Nacional del Sur. La fecha de defensa de la tesis fue el día 12 de marzo del año 2021. El desarrollo del doctorado contó con el apoyo de una beca doctoral cofinanciada del CONICET y la Secretaría de Ciencia, Tecnología e Innovación Productiva del Chubut, con lugar de trabajo en la sede Puerto Madryn de la Facultad de Ingeniería de la Universidad Nacional de la Patagonia San Juan Bosco.

1. INTRODUCCIÓN

En un Sistema de Tiempo Real (STR) los resultados deben producirse antes de un instante denominado *vencimiento*. Los STR son utilizados clásicamente en situaciones donde un fallo puede producir pérdidas económicas, daños ambientales o incluso pérdidas humanas, como sistemas de control industrial, aviónica o equipamiento médico. Por lo tanto, para garantizar la predictibilidad y determinismo de ejecución, su desarrollo emplea hardware y software de efectividad

comprobada, con diseños e implementaciones estáticas que suelen ser *ad-hoc*. En la actualidad los STR se aplican en una multitud de dispositivos como electrodomésticos, teléfonos móviles o automóviles. Estas aplicaciones contienen tareas críticas, que deben cumplir estrictamente sus vencimientos, y tareas con restricciones temporales relajadas o inexistentes. Un STR de este estilo se denomina *STR heterogéneo* (STRH), donde las tareas críticas se conocen como Tareas de Tiempo Real (TTR) y las tareas no-críticas como Tareas de No-Tiempo Real (TNTR).

Un STRH debe garantizar que se cumplan los vencimientos de las TTR y ofrecer también una calidad de servicio adecuada a las TNTR. Para esto es necesario aprovechar el tiempo ocioso que deja la ejecución de las TTR, mediante técnicas que lo calculen y administren de manera exacta en tiempo de ejecución. Este tiempo ocioso puede ser utilizado para otros objetivos como el ahorro de energía, la tolerancia a fallas o computación imprecisa. Diversos métodos han sido propuestos para este fin, muchos de los cuales subutilizan los recursos computacionales o cuentan con un costo computacional (CC) que restringe su uso en la práctica. Reducir el CC de las técnicas de administración de tiempo ocioso evita el sobredimensionamiento del hardware, con el ahorro económico correspondiente. En la actualidad, la capacidad de cómputo de los

microcontroladores permite implementar este tipo de técnicas con una sobrecarga tolerable.

En la tesis se desarrollan nuevas técnicas de evaluación de planificabilidad y de uso del tiempo ocioso, con el objetivo de que los STRH puedan ser eficaces en la planificación de sus tareas. Los métodos propuestos son implementados y ejecutados en placas de desarrollo para verificar su viabilidad práctica.

2. CONCEPTOS BÁSICOS DE STR

Según la criticidad del vencimiento, los STR se clasifican como *duros* o *críticos* (no toleran pérdidas de vencimientos), *blandos* (permiten algunas pérdidas) o *firmes* (tipifican las pérdidas según algún criterio). El cumplimiento de los vencimientos se verifica *a priori* mediante un *test de planificabilidad*. En sistemas mono-recurso (por ejemplo, con un solo CPU), el peor estado de carga ocurre cuando todas las tareas solicitan simultáneamente acceso al recurso y se lo denomina *instante crítico*. Si todas las tareas cumplen con sus vencimientos a partir de este instante entonces el STR es *planificable* [1]. Un STR se suele modelar como un conjunto de tareas periódicas [1], caracterizadas mediante un periodo (T), un vencimiento relativo (D) y un peor caso de tiempo de ejecución (C). Un STR compuesto por n tareas se describe como:

$$S(n) = \{\tau_i = (C_i, T_i, D_i) \forall 1 \leq i \leq n\}$$

Dado que las tareas son periódicas, el Factor de Utilización (FU) del STR es el porcentaje de tiempo que el recurso se encuentra ocupado y se calcula como $\sum_{i=1}^n (C_i/T_i)$. El orden de ejecución de las tareas se determina mediante un *algoritmo de planificación*, que puede ser *estático* (la planificación se realiza fuera de línea o al inicializar el sistema) o *dinámico* (la decisión se toma en tiempo de ejecución en base a prioridades asignadas a las tareas). En estos últimos las *prioridades* pueden ser *fijas* (no se modifican una vez asignadas) o *dinámicas* (pueden variar en tiempo de ejecución). El algoritmo de planificación dinámico con prioridades fijas más utilizado es Rate Monotonic (RM) [1], que asigna las prioridades de manera inversamente proporcional a los períodos. Su ejecución es predecible, es de fácil implementación y es *óptimo* entre todos los algoritmos de prioridades fijas.

3. JITTER Y PLANIFICABILIDAD

El tiempo que transcurre entre el arribo de una tarea y su activación por parte del planificador se denomina *jitter de activación*. Suponer que una tarea J arriba en el instante t , pero sufre un *jitter de activación* de j unidades. Por lo tanto, la tarea será puesta en la cola de tareas listas para ejecutar en el instante $t + j$. Si en dicho instante se activa una tarea I de mayor prioridad, la ejecución de la tarea J se pospone

hasta que la tarea I finalice. Esto genera una interferencia adicional a J , reduciendo el tiempo disponible para su ejecución. Por lo tanto, las tareas que presentan *jitter de activación* pueden no cumplir sus vencimientos a causa de la interferencia adicional de tareas de mayor prioridad. Estas situaciones deben analizarse para garantizar la planificabilidad del STR. Sin embargo, los *test* que tienen en cuenta el *jitter de activación* analizan la planificabilidad a partir de un *instante crítico* generado por el *jitter de activación*. Como no siempre existe tal *instante crítico*, STR planificables pueden ser identificados como no-planificables.

La tesis presenta un método y un algoritmo de búsqueda del *peor instante crítico con jitter* (PICJ). Con el mismo se realizan simulaciones sobre conjuntos de 10, 20 y 50 tareas, con FU entre el 10% y el 90%. Los resultados muestran que el porcentaje de sistemas con un PICJ conformado por todas las tareas decrece exponencialmente al aumentar el FU. Esto indica que considerar el *peor caso de jitter de activación* puede causar una evaluación incorrecta de la planificabilidad del STR. Dado que estos análisis son un requisito necesario en el diseño de STR críticos se propone como trabajo futuro desarrollar un método que identifique los valores reales de *jitter de activación* que pueden causar un PICJ, para utilizarlos en el análisis de planificabilidad.

4. TEST DE PLANIFICABILIDAD

En la tesis se presenta un nuevo método de evaluación de planificabilidad exacto y de bajo CC para las disciplinas de prioridades fijas RM [1] y Deadline Monotonic (DM) [2]. El método está basado en la búsqueda de un Punto Fijo (PF), técnica propuesta originalmente en [3] e independientemente en [4]. La búsqueda del PF de una tarea i utiliza la siguiente fórmula iterativa:

$$t_i^{q+1} = t_i^q + \sum_{j=1}^{i-1} (A_j^q - A_j^{q-1}) \quad \text{donde } A_j^q = \left\lceil \frac{t_i^q}{T_j} \right\rceil$$

con $t_i^{q+1} \leq D_i$, $0 \leq q \leq m$ y $1 \leq j \leq i-1$

El PF en t_i^q existe sí y sólo sí $A_j^q = A_j^{q-1}$ para todo $1 \leq j \leq i-1$. Si el primer PF se encuentra en un instante $t_i^{q+1} = t_i^q \leq D_i$, entonces la tarea es planificable y su peor caso de tiempo de respuesta es $R_i = t_i^q$. Caso contrario, la tarea no es planificable. La semilla inicial es $t_i^0 = R_{i-1} + C_i$ [5] y se actualiza durante la sumatoria como $t_i^{q+1} = t_i^q + A_j^q - A_j^{q-1}$ [6]. Si se encuentra un PF para todas las tareas en un instante menor o igual a su vencimiento, se garantiza que el STR es planificable.

Para reducir el número de invariantes a calcular, se aprovecha que el término A_j^q es una función monótona creciente a saltos, cuyo valor se mantiene constante durante el período de la tarea j . Sea entonces I_j^q el máximo

instante hasta el que A_j^q mantiene su valor. Luego, dada una iteración $q+1$, si $t_j^{q+1} \leq I_j^q$ entonces $A_j^{q+1} = A_j^q$ y no es necesario calcular A_j^{q+1} . Caso contrario se calcula y actualiza I_j^q . Esta mejora es aplicable en la evaluación de planificabilidad de cualquier política de prioridades fijas y requiere un arreglo del orden del número de tareas, que almacene los valores I_j^q .

Mediante el anterior resultado se obtuvieron los siguientes teoremas, que evitan la necesidad de una iteración adicional para verificar el PF y permiten, bajo ciertas condiciones, realizar el cálculo directo del peor caso de tiempo de respuesta:

Teorema 1 (3.1 en la Tesis): Sea un STR de n tareas. Si al finalizar la iteración q del cálculo del peor caso de tiempo de respuesta de la tarea i , se tiene que $I_j^q \geq t_j^{q+1}$ para todo $1 \leq j < i$, entonces el algoritmo iterativo ya ha encontrado un PF en t_i^{q+1} .

Teorema 2 (3.2 en la Tesis): Dada una tarea i sea $t_i^0 = R_{i-1} + C_i$. Si $I_j^0 \geq t_j^0$ para todo $1 \leq j < i$ entonces $R_i = t_i^0$ y es un PF.

Durante el cálculo del peor caso de tiempo de respuesta de una tarea i , al calcular A_j^q y actualizar I_j^q , es posible que el nuevo valor de

la semilla, $t_i^{q+} = t_i^q + A_j^q - A_j^{q-1}$, exceda el nuevo intervalo de validez, esto es $t_i^{q+} > I_j^q$. Por lo tanto, no se encontrará un PF en la siguiente iteración. Para evitar iteraciones adicionales, el cálculo del término A_j^q se reemplaza mediante el propuesto en el siguiente teorema:

Teorema 3 (3.3 en la Tesis): Dada una iteración q , el valor t_i^{q+} que maximiza el incremento de la carga de trabajo de la tarea j a partir de la semilla t_i^q , puede calcularse como:

$$t_i^{q+} = t_i^q + A_j^{q*} - A_j^{(q-1)*} \quad \text{con } A_j^{q*} = \left\lfloor \frac{t_i^q - A_j^{(q-1)*}}{T_j - C_j} \right\rfloor C_j$$

En base a estas mejoras, se desarrollaron dos nuevos métodos de evaluación de planificabilidad, RTA3 [7] (mejoras 1 y 2) y RTA4 [8] (aplica todas las mejoras). Para evaluar sus rendimientos en comparación con otros métodos (RTA [5], RTA2 [6] y HET2 [9, 10]), se realizaron simulaciones y ejecuciones en una placa de desarrollo mbed LPC1768 (ARM Cortex-M3 a 96 Mhz con 32 KiB de RAM), registrando el número de términos A_j^q requeridos para evaluar la planificabilidad de un STR y el tiempo de ejecución en microsegundos. Los nuevos métodos presentan el mejor desempeño, con un CC acotado entre $\theta(n \cdot \log(n))$ y $\theta(n^2)$. En la

Figura 1 pueden observarse los resultados para STR con 10 tareas y distribución uniforme de periodos.

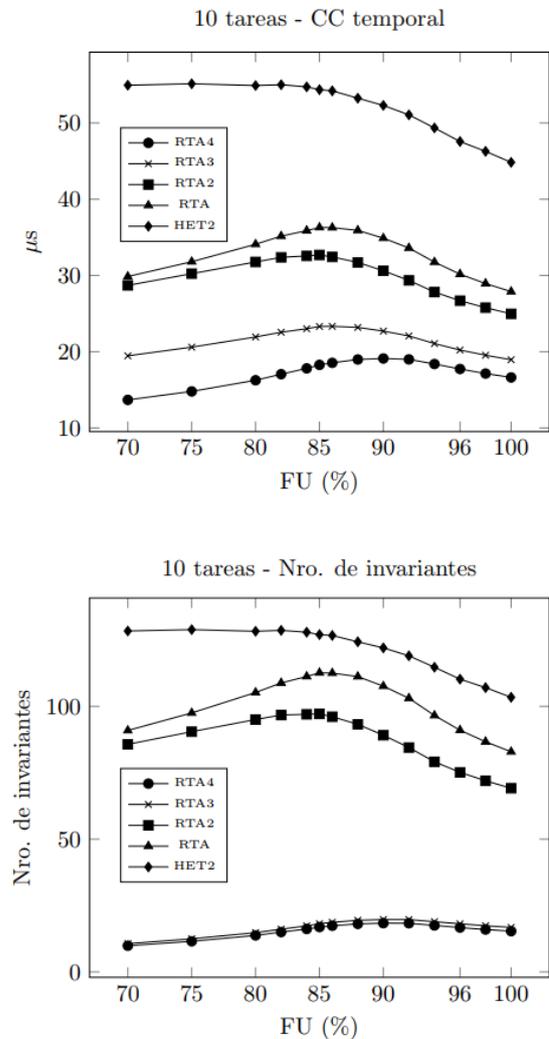


Figura 1: CC al evaluar la planificabilidad de STR de 10 tareas en microsegundos y número de invariantes calculadas.

5. MÉTODOS DE SLACK STEALING

Los métodos de Slack Stealing (SS) [11-13] permiten identificar y adelantar una parte del tiempo ocioso futuro disponible, en STR planificados mediante disciplinas de prioridades fijas como RM o DM. El tiempo ocioso que puede ser aprovechado sin

comprometer la planificabilidad del STR se denomina Slack Disponible (SD). El cálculo del SD puede realizarse fuera de línea o en tiempo de ejecución y ser aproximado o exacto. El SD puede ser utilizado para la planificación de STRH, como también para el ahorro de energía o la tolerancia a fallas.

Las primeras técnicas de SS presentan un elevado CC que imposibilita su uso en tiempo de ejecución [14, 15]. Sin embargo, aportes recientes [16-18] reducen este CC, permitiendo el uso de estos métodos en línea. Las mejoras al proceso iterativo de búsqueda de un PF presentados en la sección anterior son aplicables en los métodos de SS, reduciendo aún más el CC. El cálculo en línea del SD de una tarea i en un instante t_c se realiza de la siguiente manera [16]:

$$s_i(t_c, t) = t - t_c - \sum_{j=i}^n \left\lceil \frac{t}{T_j} \right\rceil C_j - \sum_{j=i}^n \left(\left\lceil \frac{t_c}{T_j} \right\rceil C_j - c_i(t_c) \right)$$

$$S_i(t_c) = \max_{t \in [x_i(t_c), d_i(t_c)]} s(t_c, t) \quad \text{en el menor } t.$$

Este cálculo se realiza en diversos instantes dentro de un intervalo de búsqueda, siendo el SD el mayor valor calculado en el instante t más pequeño. La mejora al proceso iterativo se aplica al primer término de la sumatoria, que es la carga de trabajo de la tarea j en el instante t (A_j^q). El costo espacial adicional es

el de un arreglo del orden del número de tareas, para almacenar los términos A_j^q .

Se modificaron los métodos de SS publicados en [17] (HeuristicSlack) y [16] (FixedSlack). Para evaluar la mejora, se contabilizó mediante simulaciones y ejecuciones en una placa de desarrollo mbed LPC1768 el número de techos y pisos requeridos para el cálculo del SD. Se evaluaron STR de 10, 20 y 50 tareas, con FU del 10% al 90% y distribuciones de periodos uniformes y por grupos. Los métodos modificados presentan un menor CC. Como resultado ilustrativo, se presenta los resultados para STR de 50 tareas con distribución de periodos por grupos (Figura 2).

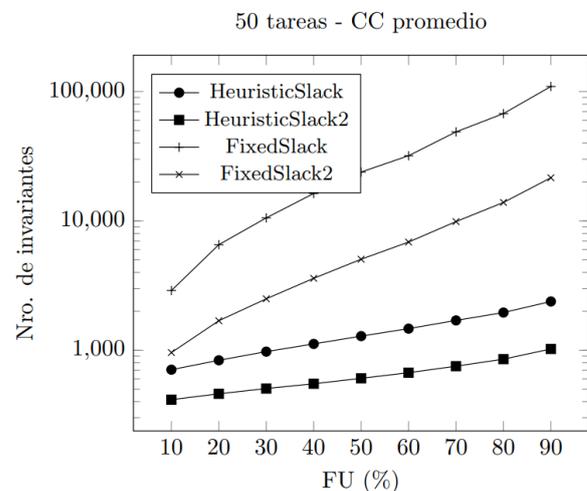


Figura 2: CC de los métodos de SS antes y después de la mejora en el cálculo de la carga de trabajo para STR de 50 tareas.

6. SLACK STEALING EN UN SOTR

Para comprobar la factibilidad de uso en tiempo de ejecución de los métodos de SS, en la tesis se presenta una implementación de

estos en el Sistema Operativo de Tiempo Real (SOTR) FreeRTOS, de código abierto y ampliamente utilizado. La implementación permite utilizar el SD para la planificación de un conjunto heterogéneo de tareas.

Se agregaron estructuras de datos para administrar los parámetros y colas de tareas adicionales (clasificación de tareas en TTR y TNTR, tareas suspendidas por falta de SD, etc.) y un API para el desarrollador. En el núcleo de FreeRTOS se modificaron las siguientes funciones:

- `vTaskDelayUntil()`: utilizada para implementar tareas con periodicidad estricta. Se agregó el cálculo del SD y la reanudación de la ejecución de TNTRs suspendidas por falta de SD.
- `xTaskIncrementTick()`: rutina de servicio de la interrupción de reloj. Se agregó la gestión de contadores de SD, la suspensión de la ejecución de las TNTRs si no existe SD y un mecanismo de control de vencimientos para las TTRs.

Por defecto FreeRTOS utiliza una política de planificación *primero en llegar primero en ser atendido* (FCFS) apropiativa por prioridades, que garantiza la ejecución de la tarea con mayor prioridad lista para ejecutar. Sin modificar este mecanismo, se altera la

composición de la cola de tareas listas para ejecutar, asignando las primeras M prioridades a las TNTR. Estas estarán presentes en la cola de tareas listas sí y sólo sí existe suficiente SD. De esta manera, las TNTR son ejecutadas si y sólo si no comprometen la planificabilidad de las TTR.

Se evaluó el CC introducido por las modificaciones midiendo el número de ciclos de CPU que requiere el cambio de contexto al finalizar la ejecución de cada TTR, donde ocurre el cálculo del SD. Las pruebas se realizaron sobre una placa mbed LPC1768. La interrupción de reloj se configuró en 1 ms, dando a cada *time slice* aproximadamente 96000 ciclos de CPU. Se empleó el método Fixed2 (ver sección anterior). Se evaluaron tres configuraciones: sin modificar FreeRTOS, calculando el SD en tiempo de inicialización y en línea. Los resultados muestran que el CC adicional de las modificaciones, sin tener en cuenta el cálculo del SD, es constante independientemente del FU. Incluyendo el cálculo en línea del SD, el costo en el peor caso no superó el 2,5% del *quantum*. En la Figura 3 se presentan estos resultados.

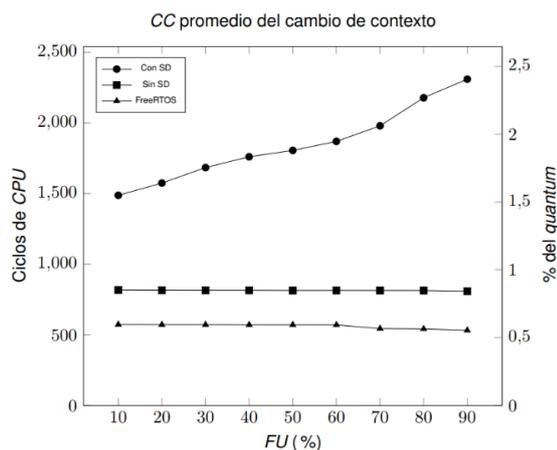


Figura 3: Comparativa del CC en ciclos de CPU del cambio de contexto de FreeRTOS.

La implementación permite verificar que la ejecución de los métodos SS propuestos es factible en un SOTR, con un CC aceptable para la ejecución de conjuntos heterogéneos de tareas. Sin embargo, podría ser utilizado para otros fines como tolerancia a fallas o ahorro de energía.

7. TAREA PLANIFICADORA

En general, los SOTR ofrecen un número limitado de políticas de planificación, como planificadores apropiativos por prioridades dinámicas. Dado que modificar el código de un SOTR no es siempre deseable o posible, se presenta en la tesis una solución mediante una *tarea planificadora* que permite implementar un planificador heterogéneo sin necesidad de modificar el SOTR. La tarea planificadora (TP) debe ejecutarse ante ciertos eventos durante la ejecución del SOTR (arriba de una tarea, finalización, bloqueo, suspensión o interrupción), en los cuales debe tomar una

decisión de planificación. El SOTR debe proveer mecanismos para reanudar y suspender tareas, cambiar sus prioridades e interceptar eventos.

La solución propuesta se organiza en dos módulos. El primer módulo se encarga del mecanismo de planificación: implementa la TP como una tarea del sistema, su activación ante los eventos correspondientes provee las estructuras de datos requeridas y la interface de programación (API) para el desarrollador. El segundo modulo implementa las políticas de planificación, por ejemplo, Dual Priority [19], EDF [1] o RM con SS. De esta manera se desacopla el mecanismo de la política implementada.

El diseño se implementó sobre FreeRTOS y se evaluó el CC ejecutando diversos STR en una placa de desarrollo mbed LPC1768. Se evaluaron STR de 10 tareas, con FU de 10% al 90%, con distribuciones de periodos uniforme y por grupos. Las pruebas muestran que el CC introducido por la TP no es excesivo, aunque debe evaluarse en cada caso particular si es aceptable. La Figura 4 presenta los resultados para periodos distribuidos uniformemente entre 25 y 1000 ms. El diseño propuesto es lo suficientemente flexible para ser adaptado a otros SOTR para implementar políticas de planificación heterogénea.

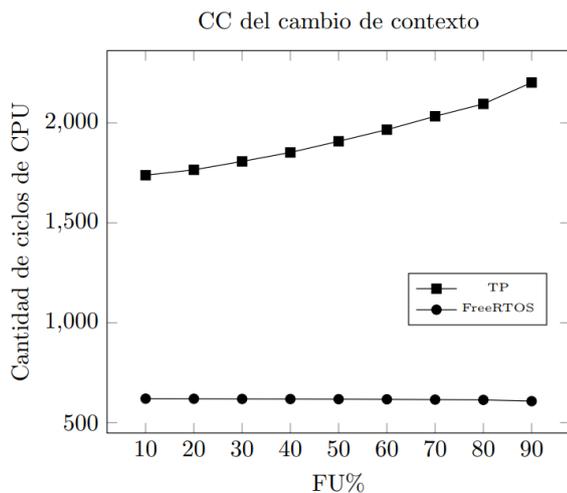


Figura 4: Comparativa del CC del cambio de contexto de FreeRTOS con y sin la TP, para STR de 10 tareas y distribución de periodos uniforme.

8. CONCLUSIONES

Los STRH requieren de métodos de administración de tiempo ocioso y de evaluación de planificabilidad para la admisión de tareas en tiempo de ejecución. Para que sea factible su empleo en la práctica, estos métodos deben presentar un CC reducido, para que la sobrecarga introducida en el sistema sea aceptable. Los métodos desarrollados en la tesis reducen el CC de los métodos exactos de evaluación de planificabilidad y de cálculo del tiempo ocioso mediante SS. Estas mejoras son verificadas mediante simulaciones, pero también mediante implementaciones en plataformas de desarrollo, cuantificando así el costo de ejecución en la práctica. Los métodos exactos de evaluación de planificabilidad de bajo costo pueden utilizarse para la aceptación en línea de nuevas tareas sin comprometer la planificabilidad del STRH, dinamizando su

composición. Aunque los métodos de SS presentados son empleados en la planificación heterogénea para brindar atención prioritaria a tareas no-críticas, estos pueden ser aplicados para el ahorro de energía o la tolerancia a fallas. De esta manera, múltiples líneas de investigación y desarrollo futuro son posibles a partir de los resultados presentados en este trabajo.

9. REFERENCIAS

- [1] C. L. Liu and J. W. Layland, "Scheduling Algorithms for Multiprogramming in a Hard Real-Time Environment," *Journal of the ACM*, vol. 20, no. 1, pp. 46-61, 1973.
- [2] J. Y. T. Leung and J. Whitehead, "On the Complexity of Fixed-Priority Scheduling of Periodic, Real Time Tasks," *Perf. Eval. (Netherlands)*, vol. 2, pp. 237-250, 1982.
- [3] M. Joseph and P. Pandya, "Finding Response Times in Real-Time System," *The Computer Journal (British Computer Society)*, vol. 29, no. 5, pp. 390-395, 1986.
- [4] N. C. Audsley, A. Burns, M. F. Richardson, and A. J. Wellings, "Hard Real-Time Scheduling: The Deadline Monotonic Approach," in *Proceedings 8th IEEE Workshop on Real-Time Operating Systems and Software*, Atlanta, GA, USA 1991.
- [5] M. Sjödin and H. Hansson, "Improved Response-Time Analysis Calculations," in *IEEE 19th Real-Time Systems Symp.*, IEEE, Ed., Dec. 1998, pp. 399-409.
- [6] J. M. Urriza, J. D. Orozco, R. Cayssials, and L. Schorb, "Reduced Computational Cost in the Calculation of Worst Case Response Time for Real Time Systems," (in English), *Journal of Computer Science & Technology*, vol. 9, no. 2, pp. 72-81, 2009.
- [7] J. M. Urriza, F. E. Paez, J. D. Orozco, and R. Casysials, "Computational Cost Reduction for Real-Time Schedulability Tests Algorithms," *IEEE Latin America Transactions*, vol. 13, no. 12, pp. 3714-

- 3723, 2015, doi: 10.1109/TLA.2015.7404899.
- [8] J. M. Urriza, F. E. Páez, M. Ferrari, R. Cayssials, and J. D. Orozco, "A New RM/DM Low Cost Schedulability Test," in *Eight Argentine Symposium and Conference on Embedded Systems*, Buenos Aires, 2017, pp. 13-18.
- [9] E. Bini, G. Buttazzo, and G. Buttazzo, "A Hyperbolic Bound for the Rate Monotonic Algorithm," *IEEE Transactions on Computer*, vol. 52, no. 7, pp. 933-942, 2003.
- [10] R. I. Davis, A. Zabus, and A. Burns, "Efficient Exact Schedulability Tests for Fixed Priority Real-Time Systems," *IEEE Transactions on Computers*, vol. 57, no. 9, pp. 1261-1276, 2008, doi: 10.1109/TC.2008.66.
- [11] S. Ramos-Thuel and J. P. Lehoczky, "Algorithms for Scheduling Hard Aperiodic Tasks in Fixed-Priority Systems using Slack Stealing," in *Real-Time Systems Symposium*, IEEE, Ed., 7-9 Dec. 1994, pp. 22-33.
- [12] J. P. Lehoczky and S. Ramos-Thuel, "An Optimal Algorithm for Scheduling Soft-Aperiodic Tasks in Fixed-Priority Preemptive Systems," in *IEEE Real-Time Systems Symposium*, Phoenix, Arizona, EUA, 1992, pp. 110-123.
- [13] S. Ramos-Thuel and J. P. Lehoczky, "On-Line Scheduling of Hard Deadline Aperiodic Tasks in Fixed-Priority Systems," in *Real-Time Systems Symposium*, 1993, pp. 160-171.
- [14] R. I. Davis, K. W. Tindell, and A. Burns, "Scheduling Slack Time in Fixed-Priority Preemptive Systems," *Proceedings of the Real Time System Symposium*, pp. 222-231, 1993.
- [15] T.-S. Tia, J. W.-S. Liu, and M. Shankar, "Algorithms and Optimality of Scheduling Soft Aperiodic Requests in Fixed Priority Preemptive Systems," *The International Journal of Time-Critical Computing Systems*, vol. 10, no. 1, pp. 23-43, January 1996.
- [16] J. M. Urriza, F. E. Páez, R. Cayssials, J. D. Orozco, and L. Schorb, "Low Cost Slack Stealing Method for RM/DM," *International Review in Computers and Software (IRECOS)*, vol. 5, no. 6, pp. 660-667, 2010.
- [17] J. M. Urriza, "Factibilidad de Sistemas de Tiempo Real con Requerimientos Heterogéneos," Doctor, Departamento de Ingeniería Eléctrica y Computadoras, Universidad Nacional del Sur, Bahía Blanca, 2008.
- [18] J. M. Urriza, R. Cayssials, and J. D. Orozco, "A Fast Slack Stealing Method for embedded Real-Time Systems," Dep. de Ing. Eléctrica y Computadoras, Universidad Nacional del Sur, Argentina., Bahía Blanca, Internal Report May 31 2005.
- [19] R. Davis and A. Wellings, "Dual priority scheduling," in *Real-Time Systems Symposium, 1995. Proceedings., 16th IEEE*, 5-7 Dec 1995 1995, pp. 100-109, doi: 10.1109/REAL.1995.495200.

TIAE – Tecnología Informática Aplicada en Educación

Clasificación automática del grado general de disfonía

Mario Alejandro García (mgarcia@frc.utn.edu.ar)

Universidad Tecnológica Nacional Facultad Regional Córdoba

1. INTRODUCCIÓN

Tesis presentada el 3/12/2021 para obtener el grado de Doctor en Ingeniería mención Sistemas de Información. Disponible en <http://hdl.handle.net/20.500.12272/5956>.

Aclaración: en este resumen se omitieron las citas y referencias por cuestiones de espacio.

En este trabajo se lleva a cabo la clasificación de la calidad vocal con técnicas de aprendizaje profundo. El objetivo de la clasificación es el grado general de disfonía (G) en escala GRBAS. Se desarrolla una red neuronal artificial que recibe audio como entrada y predice el valor de G.

1.1. Motivación

El análisis audioperceptivo es una parte principal de la rutina de evaluación clínica de pacientes con trastornos de la voz. Esta valoración repercute en la detección y seguimiento del tratamiento tanto de patologías vocales como de otras enfermedades que, si bien no tienen origen en los órganos productores de la voz, presentan como síntoma o signo el deterioro de la calidad vocal.

La necesidad de estandarizar la valoración estimuló la creación de distintos métodos y escalas de evaluación audioperceptiva, como por ejemplo GRBAS y CAPE-V. Aún con la utilización de estas escalas, la valoración de la voz tiene una gran variabilidad entre distintos profesionales de la salud (variabilidad interevaluador) y también entre distintas valoraciones del mismo evaluador (variabilidad intraevaluador). La clasificación automática de la calidad vocal podría reducir la subjetividad del análisis actual,

estandarizando la valoración entre distintos profesionales y permitiendo resultados repetibles.

1.1.1. GRBAS

La escala GRBAS consiste en la valoración de la voz a nivel de la fuente glótica a través de 5 dimensiones, grado general de disfonía (G), rugosidad (R), soplosidad (B), astenia (A) y tensión (S). Se valora en cuatro grados, desde el "0" hasta el "3". El grado "0" indica ausencia de disfonía y el "3" disfonía severa.

La debilidad de los métodos audioperceptivos reside en la subjetividad de la valoración y en la necesidad de que sea realizada por oyentes experimentados. Una alternativa es la clasificación automática.

1.1.2. Clasificación automática de la calidad vocal

La clasificación automática de la calidad vocal en las escalas de valoración audioperceptiva es un tema activo de investigación. Usualmente se toma un conjunto de medidas acústicas como características para ajustar algún modelo de aprendizaje automático.

Las redes neuronales profundas se han convertido en el estado del arte del reconocimiento de patrones, pero han sido poco aplicados a la clasificación de la calidad vocal. La tendencia de estos modelos es utilizar los datos de entrada sin procesar o con procesamiento muy básico y realizar la extracción de características en el interior de la red neuronal.

1.2. Objetivo

Esta tesis busca contribuir a la comprensión y mejora de las redes neuronales profundas aplicadas al reconocimiento de patrones relacionados con la calidad vocal, para lo cual se propuso el siguiente objetivo general: Lograr la clasificación automática del grado general de disfonía mediante una red neuronal profunda desarrollada a partir del conocimiento del dominio.

1.3. Contribuciones

Las principales contribuciones de esta tesis son:

- Un modelo de aprendizaje profundo para estimar el grado general de disfonía G .
- Redes neuronales para crear representaciones frecuenciales internas.
 - Capa STHadamard para realizar la operación de *windowing* con redes neuronales.
 - Una red neuronal para la generación del *power spectrogram* mediante el cálculo de la transformada de Fourier de término reducido con capas de convolución.
 - Una red neuronal para el cálculo del *power cepstrum*.
- Una red neuronal para el cálculo de *shimmer* sobre audio sintetizado.
- Un conjunto de operaciones de aumentación de datos para la clasificación de la calidad vocal.

2. MATERIALES Y MÉTODOS

2.1. Diseño de la red neuronal

¿Cómo diseñar una red neuronal desde cero? Partiendo desde una arquitectura determinada se pueden modificar algunos hiperparámetros y evaluar cómo afectan estos cambios sobre los resultados, pero difícilmente se pueda llegar a una solución aceptable con este

método si el punto de partida no está suficientemente cerca del objetivo. El diseño de una red neuronal profunda, por lo menos cuando se cuenta con una cantidad reducida de datos, requiere conocimiento sobre el problema a resolver.

2.1.1. Extracción de características

Para obtener la red neuronal presentada en este trabajo se partió desde un modelo inicial creado mediante la combinación de modelos más pequeños. Se diseñó cada parte para que fuera capaz de calcular una medida acústica relacionada con la calidad de la voz. Después, mediante un proceso iterativo de prueba, análisis y mejora del modelo completo, se obtuvo la red neuronal definitiva.

2.1.2. Representación frecuencial

Según lo planteado hasta el momento, la entrada de la red neuronal será la señal de audio. Existen varias representaciones posibles para la señal de audio. Los modelos neuronales de clasificación de audio (no solo en el dominio de la salud vocal) se pueden dividir en dos clases. En la primera clase están los modelos que utilizan el audio sin procesar (*raw audio*) y en la segunda los que utilizan representaciones frecuenciales, por ejemplo el espectrograma, el cepstrum o los coeficientes cepstrales de las frecuencias de mel (MFCC).

En este trabajo estudió la integración en la red neuronal, no solo la extracción de características y la clasificación, sino también la representación frecuencial (se utiliza el término "representación frecuencial" por simplicidad, aunque en algunos casos, por ejemplo en el caso del cepstrum, los datos no están en el dominio de la frecuencia). De esta forma, durante el entrenamiento se pueden ajustar los parámetros de toda la transformación. En la figura 2.1 se muestra el esquema del enfoque elegido.

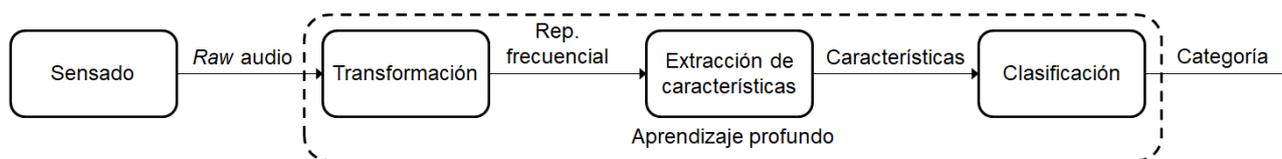


Figura 2.1: Esquema de un clasificador basado en redes neuronales que recibe raw audio como entrada.

2.2. Evaluación de resultados

La calidad de la voz es una interacción entre el estímulo acústico proveniente de la voz y un oyente; la señal acústica en sí misma no posee calidad vocal, esta se evoca en el oyente. La valoración audioperceptiva de la voz es subjetiva y se ve afectada por múltiples factores. Como consecuencia, el nivel de concordancia intraevaluador y, más aún, el nivel de concordancia interevaluador, pueden ser muy pobres.

Si suponemos que la “respuesta correcta” existe, las variaciones deben ser consideradas ruido, donde el ruido es un valor aleatorio cuya distribución depende de los datos y del evaluador. Este componente aleatorio implica que es imposible lograr una clasificación automática que concuerde exactamente con una clasificación humana particular. En consecuencia, para medir el rendimiento del clasificador, su exactitud debe ser analizada según el contexto de la base de datos (voces + valoraciones) utilizada.

En este trabajo se entrena una red neuronal utilizando como referencia la valoración de un evaluador. Después, la diferencia entre los resultados del método automático y los datos de referencia se comparan con las variaciones intraevaluador e interevaluador de expertos sobre los mismos datos. Se espera que el rendimiento del modelo desarrollado sea cercano al rendimiento humano intraevaluador en las métricas de exactitud y error absoluto medio (MAE por su sigla en inglés).

El grado G es una variable categórica ordinal, sus valores representan una jerarquía y, por lo tanto, no todos los desacuerdos tienen la misma importancia. El MAE tiene en cuenta esta diferencia.

2.3. Base de datos

Durante el desarrollo y evaluación de la red neuronal y sus partes se utilizaron tres bases de datos. La base de datos principal, usada para entrenar y evaluar el modelo final, es la *Perceptual Voice Qualities Database* (PVQD). Esta última está formada por 296 audios que contienen vocales sostenidas /a/ y su clasificación en escala GRBAS realizada dos veces por cada uno de tres especialistas clínicos.

3. REPRESENTACIÓN FRECUENCIAL DEL AUDIO

Para la definición del modelo inicial se llevó a cabo el desarrollo de tres redes neuronales aplicables a la representación frecuencial. Para cada una se realizaron pruebas con el objetivo de determinar la capacidad de representación y de adaptación de los parámetros. A continuación se mencionan de forma muy resumida los tres desarrollos.

3.1. Cálculo del espectrograma.

Se crea un modelo de red neuronal, con una capa de convolución, que calcula la transformada discreta de Fourier de tiempo reducido (STFT) para predecir la magnitud del espectrograma de la señal de entrada. Los parámetros del modelo se pueden calcular de forma directa o se pueden obtener mediante

entrenamiento con el método del gradiente descendiente.

3.2. Cálculo del cepstrograma.

Se crea una red neuronal que calcula el *power cepstrum* para distintos segmentos de tiempo. La red calcula internamente la STFT y después el cepstrum. Se realizan experimentos de entrenamiento sobre dos versiones de la red y distintas condiciones iniciales. Los resultados indican que el modelo es capaz de calcular el *power cepstrum* correctamente, pero no tiene la capacidad de adaptarse a otros problemas debido a que la derivada de la función logaritmo no es adecuada para retropropagar el error.

Un aporte importante surge del modelo alternativo analizado. La red neuronal sin el logaritmo, sí se puede entrenar para adaptarse a otros casos. Este modelo alternativo, equivalente a la función cuadrado de la autocorrelación, mantiene algunas de las ventajas del cepstrum, como por ejemplo, información para detectar la frecuencia fundamental de la señal.

3.3. Windowing con redes neuronales

Antes del cálculo del espectrograma, usualmente la señal es multiplicada por una función ventana para reducir el manchado espectral. La elección de la función ventana depende del dominio del problema. Se desarrolló una capa neuronal llamada STHadamard que permite a una red neuronal realizar la operación de *windowing* con parámetros (ventana) adaptables.

4. EXTRACCIÓN DE CARACTERÍSTICAS Y CLASIFICACIÓN

En esta sección se definen las características (calidades) del modelo inicial en relación a

la extracción de características (*features*) y la clasificación.

El diseño de la red neuronal se enfoca en tres particularidades de las voces patológicas, las cuales se agrupan en medidas de perturbación, que miden la variabilidad a corto tiempo de los ciclos glóticos con respecto a la frecuencia (*jitter*) y la amplitud (*shimmer*); y las de medidas de ruido vocal. Las tres medidas, al menos en sus métodos de cálculo, están relacionadas con la frecuencia de vibración de las cuerdas vocales, llamada frecuencia fundamental (F0).

4.1. Perturbaciones de amplitud

Shimmer es una medida acústica relativa a las perturbaciones de amplitud de una señal. Las variaciones de este tipo en la voz humana son perceptibles al oído y están asociadas a la calidad vocal

Para definir la arquitectura de una red neuronal que capture información sobre las perturbaciones en amplitud se programó generador de audio modulado en amplitud, cuyos parámetros son F0, la frecuencia de modulación (fm) y la sensibilidad del modulador (k).

Inicialmente se analizó la relación que tiene *shimmer* con F0, fm y k. Esta es una relación no lineal, pero no demasiado compleja. Se realizaron pruebas sobre redes neuronales formadas por capas densamente conectadas para aproximar *shimmer* utilizando como entrada los parámetros F0, fm y k. Como resultado de una búsqueda en rejilla (*grid search*) en el espacio de los hiperparámetros, se determinó que una red con dos capas de 20 neuronas cada una y función de activación tanh más una neurona lineal de salida, era capaz de aproximar *shimmer* con alta precisión. Sobre esta base se generaron modelos de aprendizaje profundo para problemas de aproximación de *shimmer* con

complejidad ascendente. El modelo final obtenido tiene el espectrograma como entrada y está formado por una capa de convolución, una capa de *maxpooling* y tres capas densas.

4.2. Perturbaciones de la frecuencia fundamental

La voz humana es una señal cuasi-periódica. Para que una voz produzca sensación de naturalidad deben existir pequeñas variaciones en la duración de los ciclos glóticos, pero una irregularidad excesiva en la forma de onda durante la fonación sostenida se asocia con patologías de la voz.

Las perturbaciones de F0 son evidentes en los espectrogramas y cepstrogramas. En ambos,

la perturbación se presenta como cambios de fila, algunas veces periódicos, en la posición de F0 y sus armónicos. Dado este efecto, se considera que utilizando capas de convolución, una red neuronal no debería tener dificultad en reconocer las perturbaciones de F0. Este caso es más simple que el de *shimmer*, donde las perturbaciones de amplitud se manifiestan en las magnitudes de las representaciones, pero sin provocar cambios en la posición de los máximos de energía. En la figura 4.1 se muestran los espectrogramas (izquierda de a, b y c) y el cepstrum en el tiempo (derecha de a, b y c) de tres audios de la PVQD.

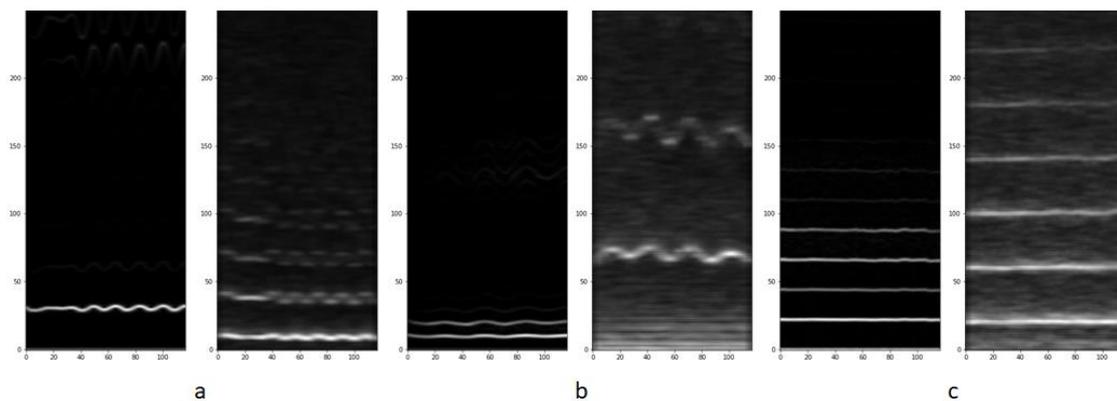


Figura 4.1: Comparación del espectrograma (izquierda) y el cepstrograma (derecha) de tres audios (a, b y c) de la PVQD.

4.3. Ruido

El ruido puede ser una manifestación del esfuerzo vocal, y al mismo tiempo, el esfuerzo vocal es un síntoma clínico frecuente en personas con patologías vocales.

Una medida relacionada con el ruido es la prominencia del pico del cepstrum (CPP por su sigla en inglés). Este se calcula como la diferencia de amplitud entre el pico cepstral (altura del mayor *rahmonic*) y el valor correspondiente en la línea de regresión directamente debajo del pico.

En este trabajo, el diseño de las capas de extracción de características permite que la

red neuronal cuente con la información necesaria para calcular un valor cercano a la *smoothed cepstral peak prominence* (CPPS), una variante del CPP. El suavizado se realiza con una capa de convolución 2D que se conecta a la entrada con un *kernel* de tamaño $n \times 1$. Los pesos del *kernel* se inicializan con los valores de una campana gaussiana. Sobre la salida de la convolución se calculan el máximo y el valor medio con capas de *maxpooling* y *averagepooling* respectivamente. Finalmente, las salidas de las capas de *pooling* se conectan a una única neurona con activación lineal.

4.4. Clasificación

La clasificación se realiza con capas densamente conectadas. La capa de salida tiene cuatro neuronas con función de activación sigmoïdal. El resto de las capas utiliza función de activación ReLU. La cantidad de capas, la cantidad de neuronas por capas y el uso de técnicas de regularización o *dropout* se definió durante el proceso de optimización del modelo.

5. MODELO INICIAL

El modelo inicial es una red neuronal creada a partir de los desarrollos de las secciones anteriores. Es el punto de partida del proceso de optimización que se lleva a cabo más adelante para obtener el clasificador del grado general de disfonía. Hay dos versiones de modelo inicial, MIA (figura 5.1) y MIB.

MIA recibe como entrada *raw audio* de un segundo de duración, calcula la operación de *windowing* (etiqueta A1 en la figura) y después el cuadrado de la autocorrelación (operaciones A2 a A8). Se extraen características por dos caminos, el camino 1 (operaciones B1 a B5) para las perturbaciones de amplitud/frecuencia y el camino 2 (C1, C2 y C3) para el ruido. La clasificación se realiza con una capa densamente conectada de cuatro neuronas (D2).

La variante MIB es una red neuronal reducida que recibe el cepstrum como entrada. La arquitectura se obtiene eliminando en MIA la capa A8 y capas anteriores. El objetivo de esta variante es determinar si, para predecir G en el contexto de este trabajo, es más importante el logaritmo en el cepstrum o la capacidad de ajustar los pesos en la etapa de representación frecuencial.

6. EXPERIMENTOS

Se realizaron experimentos sobre los modelos de clasificación iniciales MIA y MIB. El

objetivo de estos experimentos es tanto la definición del modelo final como su evaluación.

6.1. Definición del modelo final

El modelo final es el resultado de un proceso de búsqueda en el espacio de los hiperparámetros, donde se realizaron pruebas sobre cada modelo y se ensayaron modificaciones. Los hiperparámetros se modificaron en base al análisis de la evolución de ambas métricas durante el entrenamiento, la evolución y configuración final de los pesos y de las salidas intermedias (salidas de las capas ocultas).

6.2. Datos

Los datos de la PVQD, tal como indican sus autores, deben ser revisados. Además de la revisión, durante este trabajo se añadió la valoración en escala GRBAS de un profesional local.

Las tareas ejecutadas sobre los datos son la segmentación de la vocal /a/ sostenida, el renombrado de archivos, la generación de una planilla de valoración, el análisis y valoración (por parte del evaluador local), la selección y corrección de archivos de audio y la generación de base de datos transaccional. En total se eliminaron 23 audios porque no cumplían criterios de calidad y se editaron otros para eliminar ruido.

6.2.1. Aumentación de datos

Las redes neuronales profundas tienen muchos parámetros y, por lo tanto, son propensas al sobreajuste. Una forma de lidiar con el sobreajuste es aumentar la cantidad de datos de entrenamiento. En este trabajo se realizaron tres transformaciones, desplazamiento en frecuencia, segmentación por tiempo (*cropping*) y *flipping*. Los datos

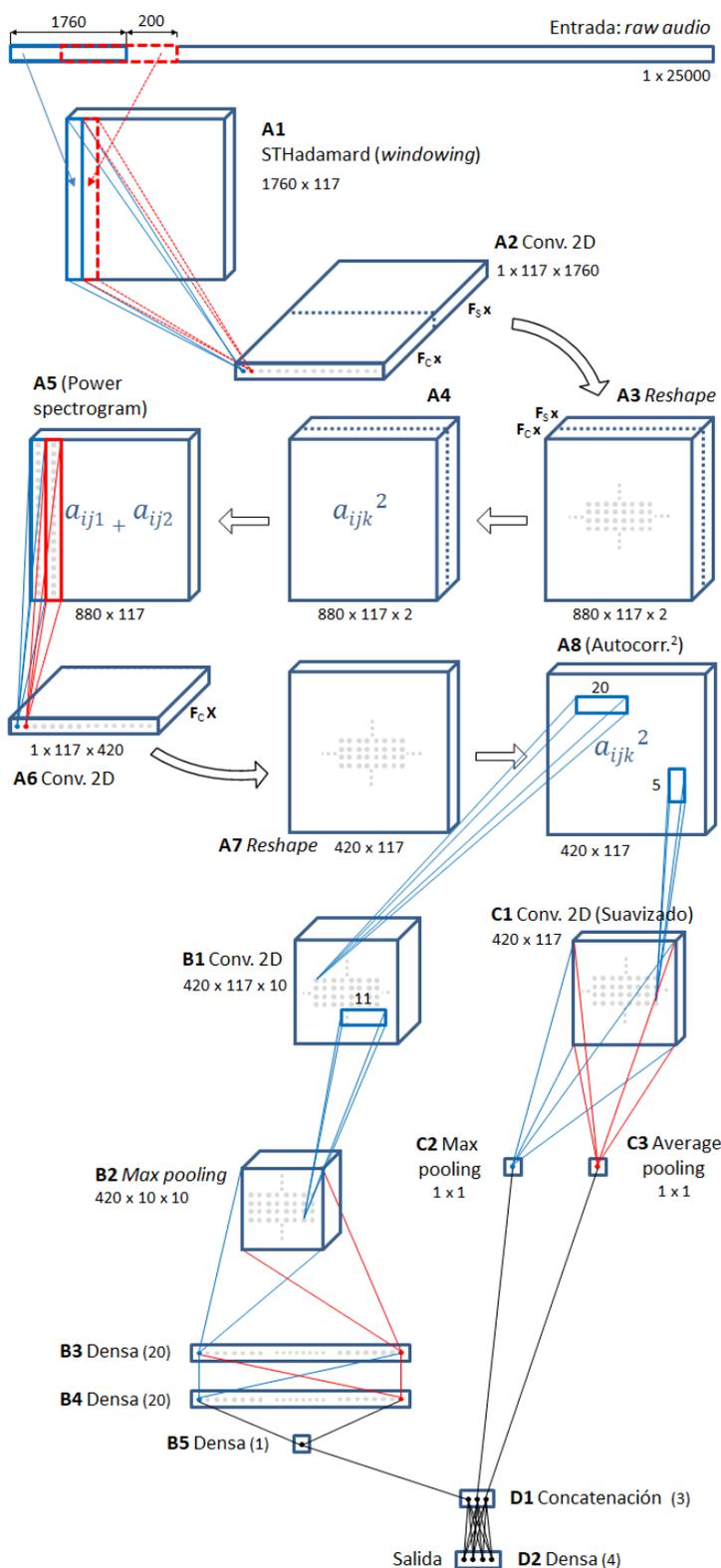


Figura 5.1: Modelo inicial MIA.

transformados se anexaron a la base de datos aumentando su tamaño.

6.2.2. Generación de los juegos de datos

Para la evaluación final del modelo (test) se reservó el 20% de los datos. Con los datos restantes se utilizó un esquema de validación

cruzada de k iteraciones (*k-fold*) para $k=4$. En total fue necesario crear 5 juegos de datos del mismo tamaño.

RESULTADOS

Durante el proceso de optimización, y a pesar de probar muchas alternativas, no se logró ninguna configuración que produjera resultados satisfactorios partiendo desde el modelo MIA. En contraste, partiendo desde MIB sí se logró predecir G. A continuación se presentan los resultados del modelo definitivo (figura 7.1) obtenido a partir de MIB.

7.1. Modelo final

Durante el proceso de optimización se observó que suavizar la entrada, no solo para el cálculo de CPPS, sino para toda la extracción de características, mejora la clasificación. También se observó que era beneficioso realizar distintos grados o escalas de suavizado, es decir más de un suavizado con ventanas de ancho distinto.

7.1.1. Extracción de características

La extracción de características, al igual que en el modelo inicial, se hace por dos caminos distintos de la red. A diferencia del modelo inicial, en el modelo final la entrada de cada camino no es el cepstrograma sino los suavizados del mismo. Tal como se puede ver en la figura 7.2, ambos caminos sufrieron cambios durante la optimización (principalmente el camino 1), pero la estructura es similar a la del modelo inicial.

7.1.2. Clasificación

En la operación D1 los dos caminos de extracción de características se unen, las salidas de ambos se reordenan y concatenan formando un vector de tamaño 48. Este vector es la entrada de una capa densa (D2) de 3 neuronas con función de activación ReLU, la que se conecta a la capa densa D3, con 10

neuronas con activación ReLU y, por último, la salida de D3 es la entrada de las cuatro neuronas de salida.

7.2. Métricas

El modelo final obtuvo una exactitud media de 0.711 y MAE 0.303. En la figura 7.2 se muestran los diagramas de caja de las dos métricas para las 10 ejecuciones del proceso. Los resultados se comparan con la concordancia y el error absoluto intraevaluador e interevaluador para los mismos datos.

En la matriz de confusión de la tabla 7.1 se comparan las valoraciones del evaluador local con las salidas del clasificador para la primera ejecución.

7. DISCUSIÓN

Una clasificación, con datos balanceados, realizada al azar sobre cuatro clases tendría un acierto del 25%, por lo tanto el 71.1% de acierto informado evidencia que la red neuronal profunda obtenida es capaz de reconocer ciertos patrones de la voz que están relacionados con la percepción de la calidad.

No solo la exactitud indica un buen rendimiento, el MAE y la matriz de confusión (tabla 7.1) muestran que, para los casos donde no hay coincidencia, el modelo asigna clases cercanas a la de referencia; tal como ocurre entre expertos humanos. Para ambas métricas, el clasificador automático logra valoraciones con variaciones muy cercanas a las que obtendría un evaluador medio tras evaluar por segunda vez el mismo conjunto de audios (variabilidad intraevaluador). También para ambas métricas se puede ver una gran mejora comparando con la concordancia y el error interevaluador. Los resultados son prometedores y permiten pensar en futuras aplicaciones clínicas de clasificadores basados en la red neuronal obtenida.

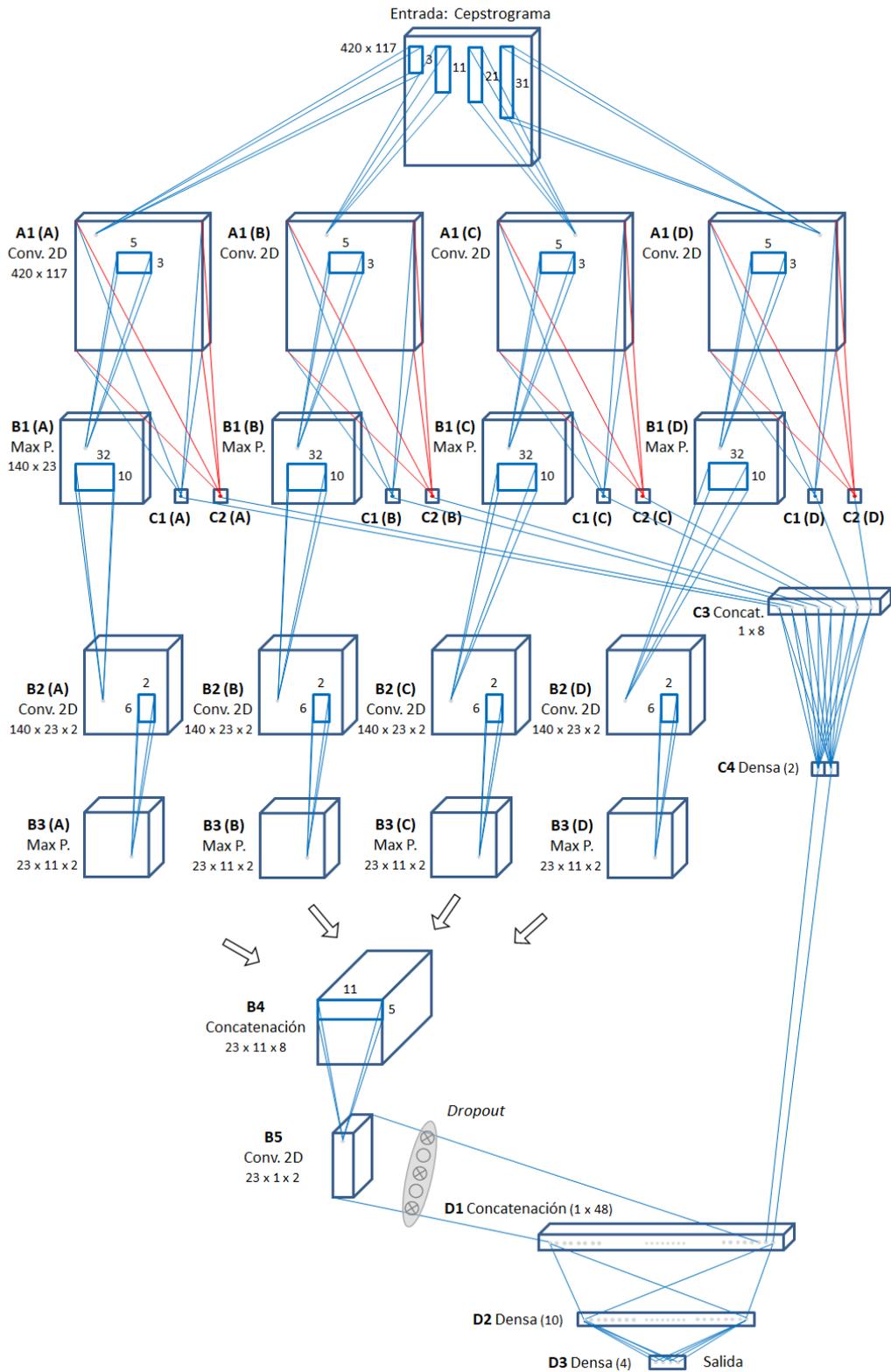


Figura 7.1: Modelo final de clasificación del grado general de disfonía.

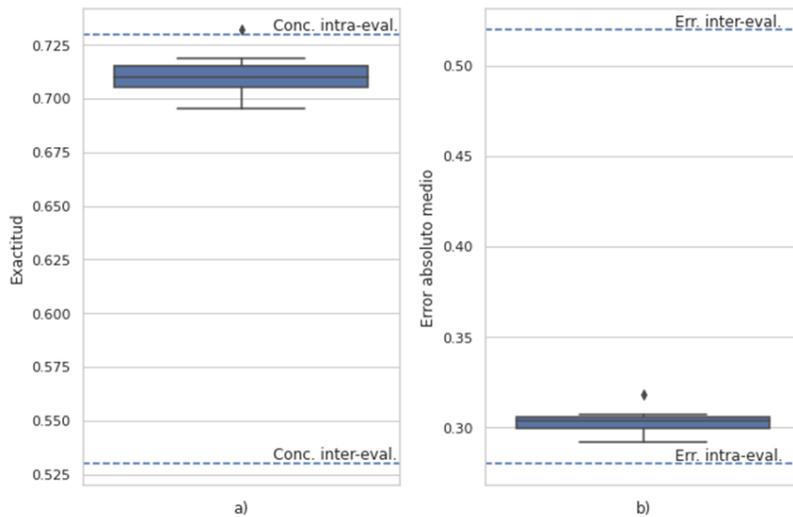


Figura 7.2: Diagrama de caja para la exactitud (a) y el MAE (b) en los resultados del experimento. En ambos casos se puede ver la comparación con los valores (concordancia y error absoluto) medios interevaluador e intraevaluador.

		Red neuronal			
		0	1	2	3
Eval. local	0	67	41	0	0
	1	14	70	24	0
	2	0	26	74	8
	3	0	2	9	97

Tabla 7.1: Matriz de confusión entre la valoración del evaluador local y la valoración automática para la primera ejecución de la evaluación.

Como parte de la discusión también se analiza la arquitectura de la red neuronal obtenida y se asocian los múltiples suavizados con la teoría del espacio escalar. Por otro lado, se plantean como limitaciones para la aplicación clínica del modelo obtenido, algunos desafíos conocidos para la aplicación de la Inteligencia Artificial en la práctica clínica. Los trabajos futuros están relacionados con los estos desafíos, que incluyen la incorporación de bases de datos multicéntricas y más grandes, lo que además permitirá probar arquitecturas más complejas.

8. CONCLUSIÓN

Se concluye que el modelo obtenido es capaz de predecir el grado general de disfonía con una exactitud cercana a la de un evaluador humano para los datos de la PVQD.

Con respecto a la arquitectura del modelo se concluye que el cepstrum no puede ser calculado por una red neuronal si los pesos de las capas anteriores deben ser ajustados y que, al menos bajo las condiciones de los experimentos realizados, una red neuronal no

puede calcular una representación interna equivalente al cepstrum que sea útil para la clasificación. En referencia a la extracción de características, se concluye que una red neuronal profunda diseñada para reconocer patrones de perturbación de amplitud, perturbación de frecuencia y ruido obtiene información útil para la predicción del grado general de disfonía.

Por último, se concluye que el modelo presentado puede ser un punto de partida para futuros desarrollos de clasificadores de la calidad vocal.

Gobernanza de Datos en los Procesos de Negocio para las Instituciones de Educación Superior

Ana Muñoz, Santiago Pérez, Maria Stefanoni, Daniela Carbonari,
Bruno Roberti, Higinio Facchini, Jimena Pérez
CeReCoN (Centro de Investigación y Desarrollo en Computación y Neuroingeniería)
Facultad Regional Mendoza, Universidad Tecnológica Nacional
Rodríguez 273, Mendoza, Argentina - 0261-5244576
(santiagocp, maria.stefanoni,dcarbonari)@frm.utn.edu.ar;

Ana Muñoz, Jimena Perez
Centro de estudios en Microelectrónica y Sistemas Distribuidos Universidad de Los Andes
Av. Alberto Carnevali, Mérida 5101, Mérida, Venezuela +58 274-2401111
(ana.munoz, jimena.perez)@ula.ve

RESUMEN

Las organizaciones requieren fuentes de datos confiables, coherentes y disponibles que soporten la toma de decisiones y el mejoramiento de sus capacidades operativas. Estas se han dado cuenta de la importancia de gestionar y gobernar sus datos como un recurso a nivel estratégico que optimicen sus procesos. "La gobernanza de datos se define como el ejercicio de autoridad y control (planificación, seguimiento y ejecución) sobre la gestión de los activos de datos". Esto implica que el Gobierno de datos cubre los estándares y principios de la gestión de datos. Una norma define requisitos, especificaciones, pautas o características para un determinado material, producto, proceso o servicio.

Existen varios modelos de gobernanza para guiar los procesos de estudio, y de implementación de Gobernanza de datos, tales como: IBM, ORACLE, el Data Governance Institute (DGI), ISO/ IEC JTC 1 / SC 4 , Collibra , DAMA DMBOK , así como los modelos emergentes de gobernanza descentralizada basados en tecnología de blockchain distribuida (Distributed data ledger: Libro de registros distribuido).

Sin embargo, existe una vacancia importante cuando la gobernanza de datos se refiere a los

procesos educativos, en si mismos, dentro de las Universidades.

Esta investigación propone el desarrollo de una ontología para el gobierno de datos, que aplicada a los procesos de enseñanza-aprendizaje en las Universidades, guíe en el uso de modelos de gobierno de datos existentes.

Los integrantes de este Proyecto tienen formación de posgrado, han dirigido tesis de posgrado y/o han formado parte de grupos y proyectos de Investigación en Educación.

Palabras clave: gobernanza, ontologías, procesos educativos, procesos universitarios

CONTEXTO

La línea de investigación está inserta en el ámbito del Centro UTN CeReCoN (Centro de Investigación y Desarrollo en Computación y Neuroingeniería), de la Facultad Regional Mendoza, de la Universidad Tecnológica Nacional, y del Centro de Microelectrónica y Sistemas Distribuidos (CEMISID) de la Universidad de los Andes.

INTRODUCCIÓN

Muchas organizaciones han puesto en marcha iniciativas de transformación digital. Cada vez hay más herramientas digitales disponibles, y muchas de ellas tienen un gran impacto en los procesos de negocio. Sin

embargo, sólo un pequeño número de organizaciones tienen sus procesos de negocio lo suficientemente controlados como para convertirse en todo el potencial empresarial de las nuevas tecnologías digitales. Las capacidades adecuadas de gestión de procesos de negocio (BPM, Business Process Management), tienen un impacto significativo en el valor logrado a través de la digitalización.

Esto es especialmente cierto para el establecimiento de un adecuado gobierno de los procesos de negocio. La gobernanza de los procesos mantiene el rumbo de los mismos. Identifica los ajustes necesarios del proceso, define las acciones requeridas y garantiza su ejecución. Esto tiene un impacto significativo en la obtención y el mantenimiento de los beneficios de la digitalización que se pretenden, así como en el rendimiento continuo de los procesos. Sin embargo, los procesos digitales tienen nuevos requisitos y las herramientas de gestión de procesos digitales ofrecen nuevas oportunidades para una gobernanza eficaz. Por lo tanto, la gobernanza de los procesos debe pasar por una transformación digital en sí misma, aprovechando las herramientas adecuadas, como la gobernanza de datos, la minería de procesos o las herramientas de modelado y simulación de procesos dinámicos. El resultado es la gobernanza de los datos de los procesos digitales, como una base importante para el éxito de la transformación digital.

La Gobernanza de Datos (GD) es parte de la gestión de información empresarial. La gestión de la información se define y se entiende como lo establece en el Cuerpo de Conocimiento de Gestión de Datos, o DMBOK [1], como sinónimo de gestión de la información. Los primeros conceptos tienen que ver con qué y en dónde ocurre la gobernanza de datos (datos como un activo). Uno de los retos más comunes de la gobernanza de datos es que el flujo y la lógica de los datos pueden no seguir la estructura de una organización. El desajuste entre la estructura organizativa y el uso de los datos puede dar lugar fácilmente a silos de datos, duplicidades, responsabilidades poco

claras y falta de control de los datos en todo su ciclo de vida.

OBJETIVOS

Objetivos generales

Desarrollar un Modelo de Gobernanza de Datos para las Instituciones de Educación Superior (IES) que permita realizar la Gestión de sus Procesos de Negocio conducida por los datos.

Objetivos específicos

1. Definir los Procesos de Negocio para las Instituciones de Educación Superior.
2. Definir el programa de Gobernanza de datos para los procesos de negocio más importantes de las Instituciones de Educación Superior, a través de un caso de estudio.
3. Definir ciclo de vida de los Datos que están relacionados con los Procesos de negocio seleccionados.
4. Simulación de los Procesos de Negocio utilizando una herramienta de Minería de Procesos de Negocio.
5. Definir el modelo de Calidad de Procesos de Negocio de las IES
6. Implementar el modelo de Gobernanza de Datos para las IES.

METODOLOGÍA

Para gestionar la integración semántica entre los diferentes estándares, se propone una arquitectura conceptual de tres niveles. La arquitectura, en la capa superior, "Top Level", debería incluir una ontología con los términos principales y sus relaciones. Cada término involucrado en este nivel está vinculado al "Nivel Medio", donde se refina en módulos. Cada módulo define una ontología de dominio para el término correspondiente del nivel superior. El nivel inferior, "Low Level", compuesto por los términos y definiciones de cada estándar, se comunica a través de reglas con el nivel intermedio y los diferentes módulos que posee cada ontología.

La idea de este proyecto es obtener un modelo que describa la gobernanza federada de datos y cómo colaborar en el intercambio de datos referidos a los estándares, permitiendo que esta colaboración se lleve a cabo de forma federada. La federación permite que la colaboración se lleve a cabo en tiempos o condiciones concretos, sin que los participantes de la federación cambien su modelo de datos. Este modelo podría convertirse en un estándar para la gobernanza de datos y su proceso de integración, a través de las ontologías.

Teniendo esto en cuenta, se establecerán los métodos y tecnologías de:

- Las reglas, procedimientos, protocolos y procesos bien definidos para desarrollar, gestionar y hacer evolucionar los procesos de negocio de las IES.
- Los marcos de gestión y coordinación: diseñar y evaluar un enfoque integral de gestión de la gobernanza de datos para alinear los objetivos de los actores, la gobernanza adecuada y los recursos. Además, estudiar cómo interactúan, se comunican y colaboran los actores.
- Los modelos de madurez: para proporcionar recomendaciones claras sobre cómo impulsar mejoras basadas en el conocimiento del nivel de madurez en el que se encuentran los procesos y los datos.

ESTADO DE AVANCE

Los investigadores o los profesionales tienden a implementar el BPM y la GD por separado. Sin embargo, los conceptos básicos de los sistemas de información reconocen que la información se elabora a partir de los datos. El BPM rastreará el flujo de información de los procesos de negocio, y la GD se convertirá en una norma, política y directriz para los datos. Tanto el BPM como la GD pretenden ayudar a las IES a cumplir sus estrategias y objetivos.

Desde investigaciones anteriores [2,3], pueden evaluarse y aplicarse dos enfoques para la implementación de la GD: el primer enfoque, la implementación de la GD, que se diseña al

principio del plan maestro de la organización, y el segundo enfoque, la implementación de la GD cuando el sistema está en marcha. Desde el desarrollo de la GD para el sistema en marcha, es necesario saber cómo funciona el sistema y cómo el proceso de negocio cumple con el flujo de información, basado en el sistema y los procesos de negocio.

Se espera que el resultado de esta investigación soporte a los profesionales y académicos de las IES en la conducción de una organización a través de datos confiables.

ESTADO DEL ARTE

La gestión de procesos de negocio (BPM) es un conjunto de métodos, técnicas y herramientas para identificar, descubrir, analizar, rediseñar, implementar y supervisar los procesos de negocio para optimizar el rendimiento de la organización [4]. El BPM consta de varios componentes, entre los que se encuentran la Instancia de Proceso, el Cliente, el Actor, el Objeto, la Actividad, el Evento, el Punto de Decisión y el Resultado. Varios estudios, que se centraron en el BPM en las Instituciones de Educación Superior (IES), afirmaron que el BPM contribuyó a adquirir el flujo de información en las IES mediante la captura de los procesos de negocio [5]. Sin embargo, no todas las IES tienen un proceso de negocio estándar. Incluso diferentes facultades de una misma institución pueden tener procesos de negocio diferentes para el mismo propósito. Los datos y la información están disponibles para la mayoría de las IES, pero a veces para el proceso de acreditación, el problema es la calidad de los datos, como los datos incompletos, la incoherencia, e incluso la ausencia de los mismos.

Para guiar el proceso de implementación del gobierno de datos, IBM, ORACLE, el Data Governance Institute (DGI) y el DAMA han definido marcos de trabajo y funciones que benefician a las organizaciones en la generación de valor y en la coordinación e integración de las áreas de negocio, así mismo, investigadores

han propuesto modelos de GD centrados en roles, la identificación de responsabilidades en la toma de decisiones y acciones con el objeto de apoyar la gestión y calidad de los datos (Khatri & Brown, 2010; Thompson, Ravindran, & Nicosia, 2015; Weber et al., 2009), y hacia el diseño de procesos y estrategias que faciliten el control de los datos en aspectos o temas específicos (Al-Ruithe, Benkhelifa, & Hameed, 2016; Fu et al., 2011). Además, se han llevado a cabo trabajos que resaltan la importancia de mejorar la gestión, la calidad, disponibilidad y seguridad de los datos (Fu et al., 2011; Guerrero & Londoño, 2016; Otto et al., 2007; Wende, 2007), otras propuestas se han derivado de marcos de gobierno de TI enlazándose con tendencias tales como la computación en la nube, Big Data o inteligencia de negocios (Al-Ruithe et al., 2016; Groß & Schill, 2012; Thomas, 2006; Weber et al., 2009). CIO Watercooler, junto con VERITAS, publicaron un libro blanco sobre los resultados de su encuesta sobre el Gobierno de Datos (DG) en 2017 [6]. En el documento, el 63% de los encuestados seleccionados del Reino Unido e Irlanda dijeron que quieren una buena calidad de datos (DQ) para que sus empresas cumplan sus estrategias y objetivos. Esta investigación también afirmaba que la GD mejoraría la calidad, la seguridad y los estándares de los datos. La GD, como práctica interrelacionada, más que aislada, ha demostrado ser una medida recomendada por la industria [7]. La GD no es un problema exclusivo del departamento de TI, sino que también afecta a otras partes de la organización, ya que incluye procedimientos, normas y flujos de información que requieren atención. Otros estudios explican el éxito de la implantación de la GD en la mejora de la calidad del servicio, y en ella influyen varios factores, como la estructura organizativa y el nivel de armonización de los procesos [8].

Los investigadores o los profesionales tienden a implementar el BPM y la GD por separado. Sin embargo, los conceptos básicos de los sistemas de información reconocen que la información se elabora a partir de los datos.

El BPM rastreará el flujo de información de los procesos de negocio, y la GD se convertirá en una norma, política y directriz para los datos. Tanto el BPM como la GD pretenden ayudar a las organizaciones a cumplir sus estrategias y objetivos. Basándonos en investigaciones anteriores [9,10], existen dos enfoques en la implementación de la GD: en primer lugar, la implementación de la GD, que se diseña al principio del plan maestro de la organización, y en segundo lugar, la implementación de la GD cuando el sistema actual está en marcha. Desde el desarrollo de la GD para el sistema en marcha, es necesario saber cómo funciona el sistema y cómo el proceso de negocio cumple con el flujo de información, basado en el sistema y los procesos de negocio.

Uno de los retos más comunes de la gobernanza de datos es que el flujo y la lógica de los datos pueden no seguir la estructura de una organización. El desajuste entre la estructura organizativa y el uso de los datos puede dar lugar fácilmente a silos de datos, duplicidades, responsabilidades poco claras y falta de control de los datos en todo su ciclo de vida. En este escenario surgen numerosos retos. Puede que no existan mecanismos establecidos para la gobernanza de los datos que permitan tratar los Procesos de Negocio con la participación de diferentes Stakeholder. Otro problema es el manejo ad hoc de los datos sin procedimientos y procesos ni infraestructura de datos segura, lo que podría dar lugar a que personas no autorizadas tuvieran acceso a elementos de datos individuales.

Teniendo en cuenta estos retos, la elección del enfoque de gobernanza de datos es fundamental, es por eso que esta investigación propone analizar el impacto del BPM en el desarrollo de la Gobernanza de Datos y la importancia de su relación para mejorar la calidad de los datos. Se espera que el resultado de esta investigación ayude a los profesionales y académicos a conducir las Instituciones de Educación Superior como una organización conducida por datos.

FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está integrado por docentes investigadores, becarios graduados y alumnos del Centro UTN CeReCoN de la Facultad Regional Mendoza de la UTN, y docentes-investigadores del CEMISID de la Universidad de Los Andes, Mérida, Venezuela. Entre estos se encuentra una tesista de Maestría, que presentará su tesis de Maestría afín a este proyecto y una tesista doctoral. Las actividades se llevan a cabo en el ámbito de las instalaciones de dichos entes, que cuentan con sus propias áreas de trabajo. Al efecto, se pone en conocimiento las propias publicaciones asociados a la temática [11,12, 13].

BIBLIOGRAFÍA Y REFERENCIAS

- [1] Cupoli, P., Earley, S., & Henderson, D. (2014). Dama-dmbok2 framework. DAMA International.
- [2] Hronza, R., & Speta, M. (2013, July). Business process center of excellence at the faculty of electrical engineering at the Czech Technical University in Prague. In 2013 IEEE 15th Conference on Business Informatics (pp. 346-349). IEEE.
- [3] Karabegovic, A., Buza, E., Omanovic, S., & Kahrovic, A. (2018, May). Adoption of BPM systems for process design in a higher education institution. In 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) (pp. 0552-0557). IEEE.
- [4] Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2013). Fundamentals of business process management (Vol. 1, p. 2). Heidelberg: Springer.
- [5] Maican, C., & Lixandriou, R. (2016). A system architecture based on open source enterprise content management systems for supporting educational institutions. *International Journal of Information Management*, 36(2), 207-214.
- [6] Wang, C. S., Lin, S. L., Chou, T. H., & Li, B. Y. (2019). An integrated data analytics process to optimize data governance of non-profit organization. *Computers in Human Behavior*, 101, 495-505.
- [7] Cheong, L. K., & Chang, V. (2007). The need for data governance: a case study. *ACIS 2007 Proceedings*, 100.
- [8] Weber, K., Otto, B., & Österle, H. (2009). One size does not fit all---a contingency approach to data governance. *Journal of Data and Information Quality (JDIQ)*, 1(1), 1-27.
- [9] Hronza, R., & Speta, M. (2013, July). Business process center of excellence at the faculty of electrical engineering at the Czech Technical University in Prague. In 2013 IEEE 15th Conference on Business Informatics (pp. 346-349). IEEE.
- [10] Karabegovic, A., Buza, E., Omanovic, S., & Kahrovic, A. (2018, May). Adoption of BPM systems for process design in a higher education institution. In 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) (pp. 0552-0557). IEEE.
- [11] Muñoz, A., Pérez, S., Stefanoni, M. & Carbonari, D. (2021, October), Creación de Escenarios de Aprendizaje por Competencias en la Educación Superior utilizando Realidad Virtual y Realidad Aumentada, ENIDI 2021, 19, 20 y 21 octubre/21.
- [12] Muñoz, A., Pérez, S., Stefanoni, M. & Carbonari, D. (2020, November), Aprendizaje inmersivo y realidad virtual: Casos de estudio en Ingeniería, ECEFI 2020, 24 y 25 noviembre/20.
- [13] Muñoz, A., Martí, L., Sanchez-Pi, N. (2021), Data Governance, a Knowledge Model Through Ontologies. In: Valencia-García R., Bucaram-Leverone M., Del Cioppo-Morstadt J., Vera-Lucio N., Jácome-Murillo E. (eds) *Technologies and Innovation. CITI 2021. Communications in Computer and Information Science*, vol 1460. Springer, Cham. https://doi.org/10.1007/978-3-030-88262-4_2

Anexando acciones en el Aprendizaje Móvil

Mg. Roberto Bertone¹, Mg. José Luis Filippi², Lic. Guillermo Lafuente³, Mg. Carlos Ballesteros⁴,
Lic. Gustavo Lafuente⁵

LIAU⁹ - Facultad de Ingeniería – UNLPam.

pbertone@ada.info.unlp.edu.ar¹

{filippi², lafuente³, balleste⁴, gustavo⁵}@ing.unlpam.edu.ar

⁹Laboratorio de Investigación de Ambientes Ubicuos

Resumen

La integración de las TICs en las instituciones educativas es un proceso que se desarrolla desde hace varias décadas en la República Argentina. Si bien existe un fuerte consenso en la universalización del acceso a las diferentes tecnologías de la información y las comunicaciones, la discusión vigente reside en la incorporación de los dispositivos móviles (teléfonos de última generación) como instrumento mediador del Proceso de Enseñanza-Aprendizaje.

En la actualidad la educación no se encuentra limitada a los entornos formales conformado por las instituciones educativas en sus diferentes niveles, se integra con procesos formativos que tienen lugar a través de los nuevos dispositivos tecnológicos de última generación, caracterizados por su miniaturización, movilidad y conectividad permanente; haciendo posible el acceso a múltiples contextos de aprendizaje virtuales. Resulta entonces imprescindible estudiar las tendencias del aprendizaje móvil en el mundo desarrollado, realizar las adaptaciones necesarias e implementarlas en el ámbito educativo.

El propósito del proyecto, de naturaleza teórico-práctico, es el de indagar las posibilidades que ofrecen los dispositivos móviles (teléfonos inteligentes, tabletas digitales) como instrumentos aplicados a diferentes situaciones de enseñanza y aprendizaje.

Bajo esa premisa se intenta llevar a cabo un trabajo experimental en la Facultad de

Ingeniería - UNLPam, desarrollando un escenario de aplicación real con la utilización de éste tipo de dispositivos.

Palabras claves: Aprendizaje Móvil. Dispositivo Móvil. Aprendizaje Colaborativo.

Contexto

Tipo de Investigación: Aplicada

Campo de Aplicación Principal: 7 1802 Computación, 7 1803 Comunicaciones.

Campos de Aplicación posibles: 13 1040 Ciencia y Tecnología, 7 4399 Otras – Educación – Tecnología Aplicada a la Educación

Institución que Coordina el Proyecto: Facultad de Ingeniería, Universidad Nacional de La Pampa.

Introducción

Con más de 5900 millones de usuarios de telefonía móvil en el mundo, los dispositivos de última generación han transformado nuestra manera de vivir. Aunque en todos los sectores de la sociedad se hace uso intensivo de esta tecnología, los educadores y los responsables de formular las políticas educativas no han aprovechado su potencial para mejorar el proceso educativo. [1]

Múltiples necesidades se presentan en el ámbito educativo en sus diferentes niveles entre los cuales se pueden mencionar: desarrollo de contenidos bibliográfico digital de producción local y regional, incentivar el uso de aplicaciones móviles que posibiliten el intercambio de información entre los actores que conforman la comunidad educativa

(videoconferencias entre docentes, estudiantes, directivos, y comunidad en general), incorporar el uso de plataformas de formación virtual abiertas, que posibiliten incorporar recursos y/o materiales didácticos en diferentes formatos, mantener un canal de comunicación fluido entre la institución educativa y el contexto local, regional, y nacional, e implantar estrategias metodológicas para alcanzar un aprendizaje cooperativo y colaborativo.

Mark Weiser [2] menciona que: “*vamos camino a ambientes ubicuos, ambientes poblados de numerosos sensores que gracias a la miniaturización de los dispositivos son invisibles al usuario y están en permanente rastreo de la actividad humana*”. Aquí radica el objetivo primario del proyecto: indagar diferentes aplicaciones tecnológicas móviles a través de las cuales se puedan ofrecer servicios que satisfagan las necesidades de los usuarios que transitan en el ámbito educativo, profundizando en el aprendizaje móvil.

Situación Actual del Problema

La aparición de dispositivos tecnológicos caracterizados por su miniaturización, gran capacidad de almacenamiento y gran velocidad de procesamiento, dieron origen a nuevos entornos en la comunicación de las personas, con una hipercomunicación caracterizada por la multimedialidad. El envío de mensajes textuales seguidos de imágenes, sonidos y videos es parte de la actividad cotidiana.

La educación influenciada por ésta realidad, acepta la necesidad de reconfigurar el proceso educativo, partiendo de algunas premisas:

a. Redefinir nuevas acciones para docentes y estudiantes.

El creciente número de dispositivos móviles y la conectividad a internet de forma permanente, permite a los estudiantes adquirir nuevos contenidos en línea y configurar sus

conocimientos a partir de intereses propios. El estudiante es artífice de su futuro, asume un papel activo, decide que aprender y cómo.

Frente a la realidad del mundo digital que rodea a los estudiantes de hoy en día, es justo que los docentes replanteen su rol en la construcción del conocimiento sobre ellos. De allí que la competencia exigida a un profesor del Siglo XXI es preciso que se enfoquen “en las necesidades de los alumnos, supervisando su búsqueda de información e intentando facilitar la búsqueda de información individual de los alumnos ya que el papel de suministrador de conocimiento ha sido superado por las TIC” [3].

Mucho se habla de cuáles serían las competencias digitales para un docente, un acercamiento a ello lo propone Núñez-Torrón Stock [4], quien plantea cinco competencias básicas para el docente en la educación, y entre ellas se pueden destacar dos:

1. Usar herramientas de trabajo en línea.
2. Utilizar dispositivos móviles en el aula.

En la mayoría de las universidades de todo el mundo se observa la incorporación de los sistemas de Gestión de Aprendizaje (LMS), para implementar entornos virtuales que permiten la gestión del aprendizaje. [5] Sin embargo algunas investigaciones indican que las motivaciones de los estudiantes son vistas como una obligación al participar en entornos de aprendizaje institucionales cerrados, mientras que la motivación es mayor cuando el proceso formativo es abierto a diferentes tecnologías en contextos informales. Estos espacios son conocidos como PLE (Personal Learning Environments) donde los estudiantes configuran su propio entorno. [6] Algunos autores ya incorporan un nuevo concepto, el de mPLE (Mobile Personal Learning Environments) como una nueva estructura de aprendizaje para la generación que hace uso de los dispositivos móviles. [7]

b. Continuidad del aprendizaje.

El aprendizaje móvil permitirá a los estudiantes dar continuidad a su formación a

lo largo de toda su vida, a partir de plataformas que posibilitan el aprendizaje móvil caracterizado por:

- Poseer gran cantidad de información.
- Ser omnipresentes.
- Disponer de gran diversidad de recursos materiales.
- Fomentar la participación colaborativa.

Actualmente los estudiantes y futuros profesionales acceden a la información a través del aprendizaje informal, han dejado de ser únicamente consumidores para ser productores de información.

En el caso particular de los estudiantes universitarios no solo deben dominar las disciplinas correspondientes a su carrera, además deben incorporar habilidades transversales como el pensamiento crítico, resolución de problemas, persistencia y trabajo colaborativo. Sin embargo, en muchos países no se están desarrollando estas habilidades. [8]

c. Grandes volúmenes de datos. Big data.

Disponer y aprovechar ésta tecnología es una obligación para todas aquellas instituciones que manipulan grandes volúmenes de datos, en nuestro caso particular la universidad.

La gran cantidad de información accesible a partir de variadas herramientas conlleva una serie de consideraciones éticas relacionadas a la propiedad de los datos y la privacidad.

Los investigadores que estudian el aprendizaje en línea, los sistemas de tutoría inteligente, los laboratorios virtuales, las simulaciones y los sistemas de gestión del aprendizaje están explorando maneras de entender y utilizar mejor la analítica del aprendizaje, a fin de mejorar la actividad del docente y seguir avanzando en la educación para todos. [9]

En éste contexto el punto de partida debe ser siempre pedagógico. No se trata de innovar por innovar. Hay que explotar el potencial que ofrecen las nuevas herramientas

tecnológicas que hacen posible el aprendizaje móvil, a partir de una planificación docente previa dentro de un marco teórico formalizado.

La herramienta en sí misma no ofrece por sí sola resultados, pero un buen docente sabrá aprovechar las ventajas que brinda para un aprendizaje móvil de calidad.

Por ello el aprendizaje móvil (mLearning) puede incorporarse como refuerzo del aprendizaje formal, ampliando la oferta educativa y la modalidad. Lo importante es innovar en el proceso educativo en su conjunto y no solo con el dispositivo tecnológico.

El entorno educativo cambia [10], la educación se presenta como la formación de los educandos en competencias, destrezas, habilidades para desempeñarse en un nuevo espacio social, el digital [11]. Nuevos escenarios educativos, los mismos actores con un nuevo rol, y la implantación de las tecnologías móviles es el desafío actual [12].

Línea de Investigación y Desarrollo

El plan de actividades corresponde al proceso de investigación aplicada, con objetivos de corto, mediano y largo plazo, y una duración prevista de cuatro años.

Primer año.

- Identificar el porcentaje de inserción de dispositivos móviles en la institución (teléfonos y/o tablets).
- Analizar la inserción del aprendizaje móvil en las universidades de todo el mundo.
- Estudiar como el aprendizaje móvil puede cerrar la brecha entre el aprendizaje formal y el informal.
- Identificar los escenarios educativos que muestren aspectos móviles en las actividades formativas del ámbito académico propio.

- Examinar herramientas tecnológicas móviles orientadas al proceso educativo de acceso libre y gratuito.

Segundo y Tercer año.

- Definir el enfoque pedagógico que posibilite un aprendizaje móvil de calidad.
- Estudiar las tecnologías disponibles para implementar acciones de aprendizaje móvil en el contexto educativo universitario.
- Definir prioridades según necesidades de inmediatez en la virtualización de las disciplinas que se han de impartir.
- Gestionar el uso de plataformas de formación virtual que posibilite una doble modalidad, presencial y mediada por las nuevas tecnologías móviles.
- Incorporar el uso de redes sociales que favorezcan la práctica educativa.
- Confeccionar objetos de aprendizaje acorde a las herramientas disponibles en la nube y a las características de los dispositivos móviles que van a operar el producto final.
- Desarrollar aplicaciones móviles a partir de las necesidades que se presenten durante el transcurso del proceso educativo virtual.
- Capacitar a la comunidad educativa en general en el desarrollo de objetos de aprendizaje a partir de los requerimientos de carácter institucional.
- Difundir los avances a toda la comunidad universitaria los progresos a medida que van transcurriendo, a través de jornadas, congresos y/o revistas científicas.
- Propiciar el intercambio de información permanente con grupos de investigación que den valor agregado a nuestra actividad profesional.

Cuarto año.

- Instituir las aplicaciones desarrolladas para dispositivos móviles que hayan alcanzado buen nivel de aceptación.
- Conformar un repositorio de objetos de aprendizajes de acceso libre.
- Registrar y difundir los resultados alcanzados con la finalidad de que se

puedan utilizar, ampliar y mejorar a través de trabajos futuros.

- Disponer de un catálogo de trabajos realizados en el sitio web del grupo de investigación GIAU (Grupo de Investigación de Ambientes Ubicuos) a partir del cual los interesados podrán acceder al material requerido.

Presentar en jornadas, congresos y/o revistas de todo el mundo los resultados alcanzados.

Resultados Obtenidos/Esperados

El proyecto da inicio a su actividad de investigación y desarrollo durante el año 2019, por lo que sus integrantes se encuentran efectuando las tareas enmarcadas en el cuarto año. A partir de la tarea realizada por los integrantes del proyecto de I+D durante el tercer año se concretó:

- a. El desarrollo de una herramienta didáctica para la enseñanza de la Arquitectura de Von Neumann. El producto alcanzado se incorporó a la plataforma de formación virtual como material educativo a utilizar durante el ciclo lectivo 2021, y se publicó en el congreso IDETEC 2020, con el nombre Herramienta Didáctica para mejorar el rendimiento académico en el aprendizaje de la Arquitectura de Von Newman. Villa María, Córdoba. 25 de junio 2021.
- b. Se configuro un método de evaluación del aprendizaje 100% virtual. El producto alcanzado se incorporó a la plataforma de formación virtual como material educativo a utilizar durante el ciclo lectivo 2021, y se publicó en la revista TEyET. Nro. 28, marzo 2021. (396-402), con el nombre Evaluación de los Aprendizajes en periodo de Pandemia.
- c. El desarrollo de nuevos objetos de aprendizaje para las materias Introducción a la Informática, Programación Procedural, Programación Web y Autómatas y Lenguajes. El producto alcanzado se incorporó a la plataforma de formación virtual como material educativo a utilizar durante el ciclo lectivo 2021, y se publicó en las Jornadas de Ciencias y Técnica de la UNLPam

aprobada por Resolución 322/2021 Consejo Superior.

Al finalizar el proyecto se espera contribuir en la incorporación de aplicaciones tecnológicas móviles a través de las cuales se puedan ofrecer servicios que satisfagan las necesidades de los usuarios que transitan en el ámbito educativo de la UNLPam y la UNLP, profundizando en el aprendizaje móvil.

Formación de Recursos Humanos

Director de Proyecto

Co-Director de Proyecto

6 Investigadores

Referencias

[1] Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2013).

El futuro del aprendizaje móvil. Implicaciones para la Planificación y la Formulación de Políticas.

<http://unesdoc.unesco.org/images/0021/002176/217638s.pdf>

[2] Weiser M. (1991), *The Computer for the Twenty-First Century*.

[3] webdelmaestrocmf.com, “Estas son las capacidades digitales de un docente moderno.” [Online]. Available: http://webdelmaestrocmf.com/portal/5-competencias-digitales-que-deben-tener-los-profesores-actuales/?utm_source=blogsterapp&utm_medium=Facebook. [Accessed: 27-Oct-2017].

[4] A. Núñez-Torrón Stock, “5 competencias digitales que deben tener los profesores actuales.” [Online]. Available: <http://www.ticbeat.com/educacion/5-competencias-digitales-profesores/>. [Accessed: 27-Oct-2017].

[5] M. P. Prendes, “Plataformas de campus virtual con herramientas de software libre: Análisis comparativo de la situación actual en las universidades españolas,” Murcia, Project Report EA-2008-0257, Jun. 2009. [Online]. Available:

http://www.um.es/campusvirtuales/informe_final_CVSL_SF.pdf

[6] F. J. García-Peñalvo, “Docencia,” in *Libro blanco univ. digital 2010*, J. Laviña Orueta and L. Mengual Pavón, Eds. Barcelona, Spain: Ariel, 2010, pp. 29-61.

[7] G. Attwell, J. Cook, and A. Ravenscroft, “Appropriating Technologies for Contextual Knowledge: Mobile Personal Learning Environments,” in *Best Practices for the Knowledge Society. Knowledge, Learning, Development and Technology for All*, M. D. Lytras, P. O. de Pablos, E. Damiani, D. Avison, A. Naeve, and D. G. Horner, Eds. Springer Berlin Heidelberg, 2009, pp. 15-25.

[8] (World Economic Forum, 2015).

[9] Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief*. US Department of Education, Office of Educational Technology, 1-57.

[10] Weiser, M. (1998) *The future of Ubiquitous Computing on Campus*. Communications of ACM, 41-1, January 1998, 41-42.

[11] Malani R., Griswold W, Simon B, (2009) *Public Digital Note-Taking in Lectures*. Ubicomp 2009.

[12] Richards M, Woodthorpe J, (2009), *Introducing TU100 "My Digital Life": Ubiquitous computing in a distance learning environment*. Ubicomp 2009.

Modelo de Evaluación de Software Educativo Libre Basado en un Método Sistemático

Estela Fritz¹; Pablo García¹; María Eva Ascheri¹; Alejandra Zangara²

¹Departamento de Matemática
Universidad Nacional de La Pampa
Av. Uruguay 151 – (6300) Santa Rosa – La Pampa – Argentina
Tel.: +54-2954-245220 – Int. 7125
[fritzem, pablogarcia, mavacheri]@exactas.unlpam.edu.ar

²Facultad de Informática
Universidad Nacional de La Plata
50 y 120 – (1900) La Plata – Argentina
Tel.: +54-221-4277270 / 4277271
alejandra.zangara@gmail.com

Resumen

El presente trabajo se enmarca en el Proyecto de Investigación: "*Propuesta de Clasificación de software libre utilizado en la enseñanza de la programación*", que se desarrolla en el ámbito de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de La Pampa y que fue acreditado por Resolución N° 160/2018 del Consejo Directivo.

Su principal objetivo es construir un modelo para evaluar cualitativa y cuantitativamente cualquier software educativo utilizado en la enseñanza de la programación y poder incluir en ese modelo también, las características deseables de un software libre. Por otra parte, es importante mencionar que se pretende una evaluación desde el punto de vista pedagógico más que desde el punto de vista técnico, o desde la Ingeniería del software. La construcción de dicho modelo se basa en el método LSP (*Logic Scoring Preferences*)

Palabras clave: Método LSP, evaluación de software, software educativo, software libre, evaluación cuali-cuantitativa.

Contexto

La Directora del proyecto es la Mg. María Eva Ascheri de la Facultad de Ciencias

Exactas y Naturales de la Universidad Nacional de La Pampa (UNLPam) y la Co-Directora es la Dra. Alejandra Zangara de la Facultad de Informática de la Universidad Nacional de La Plata (UNLP). El mencionado proyecto incluye al Mg Ing. Pablo Marcelo García y a la Profesora Estela Marisa Fritz como investigadores. Esta última elabora su Trabajo Final para alcanzar el grado de Especialista en Tecnología Informática Aplicada en Educación en el ámbito del presente proyecto. Además, será la base para desarrollar su futura tesis de maestría a llevarse a cabo en la UNLP.

La incorporación del Mg. García se produce en 2022 y se relaciona con su larga experiencia en el ámbito de la programación (está a cargo, desde 2000, de la asignatura "Lenguajes de Programación" de la carrera "Profesor Universitario en Computación" de la FCEyN – UNLPam). De la misma manera ha trabajado en proyectos de investigación relacionados con el software libre y ha realizado diversas publicaciones en ese ámbito.

1. Introducción

Los rápidos avances en la Tecnología de la Información y la Comunicación y las cuestiones propias de esta época de pandemia, que obligaron a adoptar la virtualidad en la

enseñanza como una modalidad casi excluyente, han incrementado la relevancia de seleccionar software educativo apropiado para poder sostener y enriquecer la labor docente en el ámbito académico, en un espacio donde hasta hace algún tiempo, la presencialidad como modalidad en la enseñanza tenía un rol importante.

Los criterios para la evaluación del software educativo deben ser considerados en una instancia previa al proceso de selección del mismo. Identificar apropiadamente esos criterios es la clave para la evaluación exitosa y posterior selección del software. Por ello este trabajo propone la construcción de un modelo elaborado a través de un método sistemático: *el método LSP*

En este modelo, el docente puede ponderar cada una de las características deseables para el software, de acuerdo con la importancia o relevancia que posea dicha característica en el sistema a seleccionar.

Desafortunadamente, no hay una lista genérica de criterios que puedan ser usados para evaluar cualquier software.

Existen en la bibliografía numerosos modelos para evaluar software educativo. Algunos con énfasis en el proceso de diseño y desarrollo tal como se propone en [1]. Otros enfoques se rigen por los principios propios de la Ingeniería del Software [12]. Se busca en otros casos la calidad del software según los estándares internacionales [3] y [4]. Muchas publicaciones presentan un modelo basado en listas de criterios que se formulan en forma de preguntas. Son ampliamente difundidas en este sentido las listas de control (*check-lists*), aunque contrariamente a lo que propone el método del presente trabajo, la técnica de las listas de control no permite ponderar por relevancia los diferentes ítems. Squires y Mc Dougall [5], proponen un modelo basado en interacciones entre los actores intervinientes desde la etapa de diseño del software hasta su utilización: interacciones entre diseñadores, estudiantes y profesores tomados de dos en dos; algunas de esas interacciones tienen sentido bidireccional. Este modelo propone un

enfoque cualitativo y aporta un modo de pensar acerca de la selección del software.

Sin embargo, cuando evaluamos y seleccionamos software educativo, hay otros aspectos relacionados con el proceso de aprendizaje, que deben ser considerados, a saber:

1. El rol desempeñado por el docente en el contexto de uso del software.
2. Contribución del software educativo al desarrollo de habilidades intelectuales, estrategias cognitivas, información verbal, actitudes y otras.
3. La aproximación a los principios de enseñanza y aprendizaje subyacentes en el sistema de software.

Es por eso que los criterios técnicos ([3], [4]) son insuficientes para determinar la aptitud del software, y los criterios educacionales también deben ser considerados ([2], [5]). Ambas clases de criterios están relacionadas.

1.1 Algunos Aspectos del Diseño Instruccional

Antes de hacer una reseña del método LSP, se consideran algunos tópicos sobre principios de diseño instruccional [6]. Se mencionan aquellos considerados relevantes para establecer criterios elementales y complejos para LSP. En principio, la identificación de los objetivos o metainstruccionales hasta el proceso de evaluación de estos. Existen, según Gagné [6] cinco categorías principales de resultados del aprendizaje, tales como habilidades intelectuales, estrategias cognitivas, información verbal, actitudes y habilidades motoras. La medida en la que el software objeto de la evaluación contribuya a desarrollar alguna o algunas de estas categorías de resultados del aprendizaje, es una característica para considerar como parte de la evaluación. Una mención especial merecen las condiciones de aprendizaje necesarias para la adquisición de esas capacidades. Las diferencias individuales entre los sujetos que aprenden y cómo esas diferencias afectan a la planificación de los

escenarios de aprendizaje, es un t3pico para incluir entre los criterios considerados cuando se eval3a un software.

Tambi3n deben considerarse criterios relativos a las posibilidades de un software de incluir procedimientos de evaluaci3n.

El aprendizaje es un proceso y como tal, provoca cambios en las capacidades del sujeto que aprende. Por eso, es deseable que un software educativo promueva dichos cambios.

Cuando se examina cuidadosamente, la situaci3n de aprendizaje tiene dos partes: una externa al sujeto y una interna al sujeto que aprende. Esa parte interna deriva de aquello que el sujeto puede recordar o recuperar de su memoria. Esas partes (la externa y la interna) est3n relacionadas. Esa relaci3n existente se denomina *condiciones del aprendizaje*.

1.2 Modelo Cuantitativo

Los modelos cuantitativos buscan objetividad en la evaluaci3n. Para ello utilizan un conjunto de criterios preestablecidos relevantes y coherentes en contraste con evaluaciones subjetivas basadas en puntos de vista particulares (el del docente, el del estudiante, el del experto). Estos criterios preestablecidos valoran diferentes componentes del software, caracter3sticas deseables relacionadas con aspectos pedag3gicos y did3cticos: adecuaci3n a los objetivos de aprendizaje, enfoque pedag3gico (constructivista o cognitivista), orientaci3n hacia los alumnos, grado de intervenci3n docente, inclusi3n de m3dulos de evaluaci3n y seguimiento, entre otros.

Tambi3n es posible incluir en el modelo cuantitativo, variables asociadas con las cuatro libertades de un software para ser considerado software libre. Asimismo, ser3a posible incluir variables (o criterios) relacionados con aspectos t3cnicos de dise1o de software: accesibilidad, adaptabilidad, interoperabilidad, reusabilidad, por nombrar algunos.

Este modelo supone una evaluaci3n previa del software y los criterios

seleccionados no est3n relacionados con la utilizaci3n de 3ste por parte de los estudiantes, ya que el proceso de evaluaci3n tiene como finalidad la selecci3n del software m3s adecuado y que re3na las caracter3sticas buscadas o se aproxime a ello, lo m3s posible. Sin embargo, el modelo es muy flexible y podr3an incluirse variables (criterios) relacionadas con el uso del software una vez que el estudiante interact3a con el mismo.

1.2.1 Construcci3n del Modelo

Se describe a continuaci3n, en forma muy general, el m3todo en el que se basa el modelo cuantitativo propuesto en este trabajo: El m3todo *LSP (Logic Scoring Preferences)* [8].

LSP es un m3todo cuantitativo basado en el empleo de la l3gica continua que posibilita la creaci3n de funciones complejas que intervienen en la evaluaci3n, comparaci3n y posterior selecci3n de sistemas de software de prop3sito general [9].

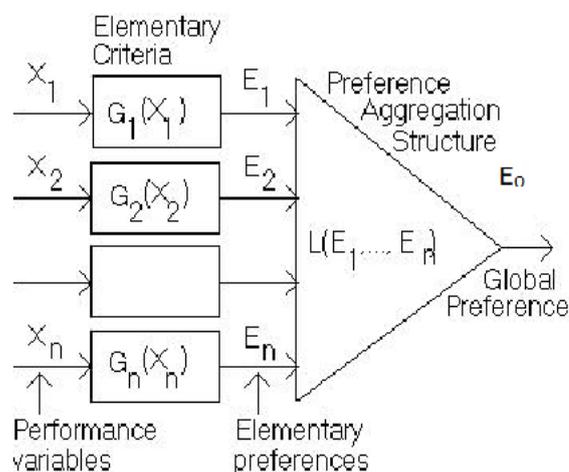


Figura 1. El proceso de evaluaci3n de LSP

Una primera etapa en la construcci3n de esta funci3n compleja consiste en establecer las caracter3sticas deseables para el software a evaluar. Esas caracter3sticas se transforman seg3n LSP en *variables de performance* (X_1, \dots, X_n). En el caso del software educativo, algunas variables de

performance que se pueden describir son las siguientes: en qué medida es tenido en cuenta el aprendizaje individual, cuantificar fases de aprendizaje inmediato y de largo alcance, gradualidad del software para el aprendizaje (¿Es posible identificar esa gradualidad?), en qué medida el software está basado en el conocimiento de cómo aprende el ser humano, en qué grado contribuye el software a los principios de contigüidad, repetición y refuerzo, qué contenidos de la memoria o capacidades aprendidas el software contribuye a recuperar, qué categorías (en cuanto a resultados del aprendizaje) el software tiende a desarrollar, cuál es el rol del docente en el proceso, en qué medida el sistema guía al sujeto hacia el autoaprendizaje. Esto por mencionar algunas.

Para cada variable de performance es necesario definir un rango aceptable de valores. Los valores de dichas variables, se transforman, por la aplicación de funciones llamadas *criterios elementales* G_i , $1 \leq i \leq n$, en *preferencias elementales* E_i , $1 \leq i \leq n$. Esos criterios elementales transforman el valor de una variable de performance en un valor perteneciente al intervalo $[0,100]$ o $[0,1]$, es decir, $0 \leq E_i \leq 100\%$ y expresan el grado de cumplimiento con un requisito del sistema que está siendo evaluado. Estas preferencias elementales van siendo agregadas mediante operadores de la lógica continua en estructuras de agregación más complejas que permiten obtener un único valor final denominado *Preferencia global E* (E_0 en la fig. 1) [10]. Existe la posibilidad de ponderar las preferencias elementales dentro de la estructura de agregación, mediante pesos W_i , que son coeficientes que multiplican a las preferencias elementales y les otorgan mayor o menor relevancia en la E . La preferencia global se interpreta como el grado global de satisfacción de todos los requerimientos especificados.

2. Líneas de Investigación y Desarrollo

El proyecto de Investigación antes citado aborda los siguientes aspectos relacionados a:

- Relevamiento de software libre utilizado en la enseñanza de la programación. Se prevé incluir aquí varias herramientas que el Mg. García utiliza en la enseñanza de los cuatro paradigmas de programación.
- Construcción de un modelo para la evaluación y selección del software libre, a partir de un método sistemático y basándose en criterios, fundamentalmente pedagógicos.
- Evaluación del software relevado con el fin de clasificarlo según los criterios preestablecidos.

3. Resultados Obtenidos/Esperados

Después de considerar los tópicos explicados anteriormente, como resultado, puede decirse que el método *LSP* propuesto constituye un muy buen método para construir a partir de él, un modelo cuantitativo de evaluación que incrementa la eficiencia de esta y la posterior selección de sistemas de software en general y software educativo en particular con filosofía GNU. Representa un método más ventajoso cuando se lo compara con otros métodos cuantitativos [10]. *LSP* tiene principalmente las siguientes ventajas relativas a los métodos cuantitativos para evaluación de sistemas:

- (1) La especificación de requerimientos es sistemática, flexible y completa.
- (2) El proceso de evaluación es sistemático y racional, el cual refleja explícita y cuantitativamente el nivel global de satisfacción de los requerimientos del usuario.
- (3) Es un modelo que permite objetividad en la evaluación.

El método *LSP* puede ser usado para la evaluación y comparación de una amplia variedad de sistemas complejos [11]. El principal poder de la estrategia de *LSP* es la capacidad de construir un modelo versátil de un conjunto de preferencias. Mediante la combinación apropiada de operadores

(lógicos) de agregación con el conjunto de preferencias, es posible derivar criterios complejos que tienen poder expresivo y flexibilidad.

El software educativo debe satisfacer una variedad de requerimientos. El nivel global de satisfacción de esos requerimientos (E) es usado para su evaluación y comparación.

4. Formación de Recursos Humanos

En el marco del presente proyecto, como ya se mencionó anteriormente, se espera que la Profesora Fritz alcance el grado de Especialista en Tecnología Informática Aplicada en Educación y que, en un futuro, pueda desarrollar su tesis de maestría en la UNLP.

Actualmente, la Profesora Fritz ha completado la etapa de búsqueda y análisis de criterios relevantes para la posterior construcción del modelo de evaluación del software principalmente, tema en el que no era experta. Esto constituye un insumo necesario para los nuevos avances en el proyecto. Se espera que durante el año 2022 se realice un fuerte trabajo de evaluación de múltiples herramientas de software libre que se utilizan en la cátedra a cargo del Mg. García.

5. Bibliografía

- [1] **Zaragoza, J., Casado, A.** (1992) “Aspectos técnicos y pedagógicos del ordenador en la escuela”. Editorial Bruno, Madrid, España.
- [2] **Gorga, G., Madoz, M., Pesado P.** (2000) “Hacia una propuesta de métrica para la evaluación de Software Educativo”, CACIC.
En línea: <http://hdl.handle.net/10915/23514>
- [3] **International Standard. ISO/IEC 9126(1991)** Information Technology – Software Product Evaluation – Quality characteristics and guidelines for their use.
- [4] **International Standard ISO/IEC 9001(1991)** Quality Systems – Model for quality assurance in design/development production, installation and servicing.
- [5] **Squires, D., McDougall, A.** (1997) “Cómo elegir y utilizar software educativo”. Traducción Pablo Manzano. Ediciones Morata y Fundación Pandeia, Madrid, 1° edición.
- [6] **Gagné, R.M., Briggs, L., Wager, W.** (1992) Ed. Harcourt Brace College Publishers, 4th edition.
- [7] **Stallman, R.** (2004) “Software libre para una Sociedad libre”. Traficantes de Sueños, Madrid.
- [8] **Dujmovic, J.J. and Elnicki, R.** (1982) “A DMS Cost/Benefit Decision Model: Mathematical Models for Data Management System Evaluation, Comparison, and Selection”. National Bureau of Standards, Washington. DC., No NBS-GCR, NTIS No PB82- 170150 (155 pages).
- [9] **Dujmovic, J.J.** (1996) “A Method for Evaluation and Selection of Complex Hardware and Software Systems”. Proceedings of the 22nd International Conference for the Resource Management and Performance Evaluation of Enterprise CS, CMG 96, 1, 368-378.
- [10] **Daso, A. et al** (2013) “Desarrollo de Modelos de Evaluación Usando Operadores de una Lógica Continua”. XV WICC (Workshop de Investigadores en Ciencias de la Computación), págs. 420-424.
- [11] **Dujmović, J. J., Hajime Nagashima.** (2006) “LSP Method and Its Use for Evaluation of Java IDEs.” International Journal of Approximate Reasoning 41.1: 3–22. Web.
- [12] **Galvis, A.** (1995). “Fundamentos de tecnología educativa”. Editorial de la Universidad Estatal a Distancia. San José, Costa Rica.

Aplicación de Analítica de Datos en espacios de Educación Superior

Gustavo Illescas¹, Elías Todorovich¹, Claudio Aciti¹, Guillermo Rodríguez², Pia Silvestrini³

¹Instituto de Investigación en Tecnología Informática Avanzada (INTIA-Centro Asociado CIC, Facultad de Ciencias Exactas, Universidad Nacional del Centro de la Provincia de Buenos Aires (UNCPBA).

²Instituto Superior de Ingeniería de Software Tandil (ISISTAN-CONICET), Facultad de Ciencias Exactas, UNCPBA.

³Instituto de Investigación en Ciencias de la Salud (IICS), Facultad de Ciencias de la Salud, UNCPBA.

Paraje Arroyo Seco, Tandil (B7001BBO), Argentina.

{illescas, etodorov, caciti, grodri}@exa.unicen.edu.ar, pia.silvestrini@salud.unicen.edu.ar

RESUMEN

El objetivo de esta investigación es realizar una Analítica de datos para la toma de decisiones sobre la información disponible en bases de datos de educación superior, en particular la información vinculada a la unidad académica donde se radica este proyecto (Facultad de Cs. Exactas, UNCPBA). El foco principal de análisis estará puesto sobre los procesos de ingreso, permanencia y egreso, como así también en el estudio y aplicación de las técnicas subyacentes dentro de las aplicaciones de *Big Data* y *Machine Learning*. Las instituciones de educación superior por naturaleza son generadoras de grandes volúmenes de datos que por lo general son de uso administrativo. Sin embargo, esas bases de datos contienen una riqueza de información para la toma de decisiones que proponemos explorar y aprovechar según lo expresado en nuestra propuesta.

Palabras clave: *Analítica de Datos, Educación Superior, Indicadores, Optimización*

CONTEXTO

Esta investigación se enmarca dentro del proyecto de incentivos 03/C314 titulado “Gestión Informática del Conocimiento como soporte para la toma de decisiones Organizacionales” y se desarrolla en el Instituto de Tecnología Informática Avanzada (INTIA) de la Facultad de Ciencias Exactas

(EXA), Universidad Nacional del Centro de la provincia de Buenos Aires (UNCPBA). Colaboran con el proyecto miembros de Universidades extranjeras:

-Laboratorio en Ingeniería de Software de la Universidad Carlos III (UC3M), Madrid, España.

-Universidad de Celaya, México.

-Universidad Autónoma de Coahuila, México.

1. INTRODUCCIÓN

Las Bases de Datos de Educación Superior (BDESUP), en particular la de EXA-UNCPBA, como lo muestra la Figura 1, están conformadas por un gran volumen de datos provenientes de los diferentes sistemas de gestión como lo son, entre otros, el sistema de alumnos (Siu-Guaraní), el de evaluación docente (Kune), el de planta funcional (Majen), el de encuestas (Siu-Kolla) (Comunidad SIU), el sistema de indicadores y el Aula virtual Moodle.

Por otro lado, existen numerosas actividades que se registran en bases de datos ad-hoc o archivos independientes como consecuencia de la falta de herramientas de software adecuadas a esos fines y que luego son de difícil interpretación a una escala global (cuando se intenta generalizar), con lo que sirven como soporte para situaciones específicas o personalizadas.

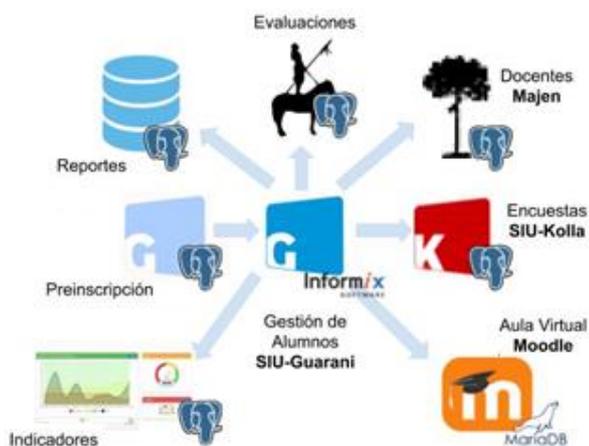


Figura 1. Esquema de Sistemas y Bases de datos disponibles en EXA. Fuente: área TICs EXA

En cuanto a los sistemas institucionales, existe la tendencia a reflejar la información a partir de la construcción de listados (también llamados reportes o informes) que son utilizados por los administrativos y equipos de gestión para realizar consultas (o filtrados) sobre el conjunto de datos disponibles y que de cierta forma son las herramientas que también se utilizan para la toma de decisiones en las unidades académicas (UA). No obstante, las BDESup contienen abundante información que no se refleja en estos listados y que generalmente no son aprovechadas para la generación de nuevo conocimiento. Por eso, la propuesta es utilizar un conjunto de técnicas y herramientas que permitan explorar todos los datos y detectar tanto información oculta como así también patrones en los datos que asistan a los tomadores de decisiones: los miembros de equipos de gestión, secretarías, directores de departamentos, coordinadores de carreras, entre otros actores de las UA, integrando así investigación con procesos de gestión.

Dadas las numerosas técnicas existentes, en una primera instancia, proponemos indagar sobre un conjunto limitado. Esto es, por un lado, las provistas por la minería de datos: patrones ocultos (Miller T. 2015; Siegel E. 2016), reglas de asociación (Concessao R. 2016), procesamiento analítico en línea (Hernández García C. et al 2016), *clustering* (Miranda M. et al 2017); y por otro, las técnicas de la optimización de la información provenientes de la rama de la estadística (Papoulis A. et al 2002) y de la investigación

de operaciones: Redes Bayesianas (Hernández Leal P. F. 2011; Oviedo B. et al 2015; Vegega C. et al 218), Cadenas de Markov (Bañuelos Saucedo A. L., 2013; Eugenia, Á. R. 2000), Procesos Jerárquicos Analíticos (Saaty, T. 2008), Árboles de decisión (Gutiérrez Pizarro C. 2014; Morrison D. 2007).

Por otro lado, a partir de la crisis sanitaria, los espacios educativos se han visto en la necesidad de utilizar plataformas de aprendizaje para interactuar activamente con sus estudiantes. Entornos educativos, tales como Moodle, se han masificado a tal punto de ser el principal medio de desarrollo académico. Esta superpoblación de uso ha expuesto una falta de evolución en el sector, el cual desaprovecha las nuevas tecnologías interactivas. En este sentido se propone realizar un asistente virtual (Carlos et al 2021; Chinedu, Ade-Ibijola 2021) que complemente activamente la utilización de la plataforma Moodle, contando con diversas funciones adecuadas al rol de quien lo utilice en la plataforma. El proyecto presentado a una beca EV-CIN se denomina Virtual Moodle Bot (VMB): Un Bot virtual en Moodle para asistir y complementar actividades académicas en espacios de educación superior.

Metodología

Todo proyecto de investigación conlleva un profundo proceso creativo para generar ideas que den solución a las preguntas de investigación que lo motivan, es por ello, que para el desarrollo de este proyecto se propone utilizar el método de trabajo conocido como *Design Thinking* (Brown T. 2009), el cual, es un método de trabajo en equipo que propicia la colaboración y la entrega frecuente de resultados a través de varias iteraciones.

Cuando se desarrolla un proyecto de carácter interdisciplinario, como el que se presenta en esta propuesta (investigadores de distintos institutos, miembros de equipos de gestión, no docentes de distintas secretarías), puede llegar a ser un reto aprovechar el talento que cada persona puede aportar desde su área de experiencia. Este reto, es otro motivo para emplear esta metodología como marco de trabajo, ya que promueve la colaboración y a

aportación de ideas a través de los siguientes principios:

- a) Definición clara de roles de trabajo.
- b) Delimitación de fases de trabajo.
- c) Definición de pautas de comportamiento.
- d) Diseño de soluciones centrado en el usuario (o la persona).
- e) Desarrollo iterativo e incremental.

Si bien el método fue diseñado en el ámbito del diseño industrial, en la actualidad, es un marco de trabajo que también se está empleando en el desarrollo de proyectos de investigación y en el desarrollo de programas académicos debido a los resultados demostrados en la generación de nuevas ideas e innovaciones.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El proyecto de incentivos que contiene esta línea de investigación comenzó a desarrollarse a mediados del 2021 luego de recibir la aprobación del organismo correspondiente. El estado de avance está relacionado con el estado actual del conocimiento sobre el tema donde los integrantes y colaboradores han realizado experiencias en la temática que han sido reportadas en diferentes comunicaciones desde 2014 a la fecha. A partir de estas experiencias es que surge la iniciativa para formular la propuesta e integrarse como equipo de trabajo. El foco principal de análisis estará puesto sobre los procesos de ingreso, permanencia y egreso siendo los dos últimos los que mayor cantidad de datos poseen dado que registran toda la actividad estudiantil. Sobre esto último, una gran ventaja es la constitución reciente de áreas que se ocupan de estas problemáticas en varias UA de la UNCPBA, lo que las convierte en las principales beneficiarias de la investigación propuesta.

Los antecedentes del equipo de trabajo están vinculados con las experiencias e investigaciones personales que los integrantes y colaboradores han desarrollado vinculadas a la temática propuesta.

-El Dr. Gustavo Illescas ha participado en actividades de EXA en el área de Coordinación de Asuntos Estudiantiles y Graduados (CAEG), en la Sec. General y en la Sec. de

Extensión. Rol: coordinador del proyecto de implantación de un sistema de indicadores (2018-actualidad).

-El Dr. Todorovich ha participado como director en el Proyecto de Mejoramiento de la Enseñanza en Carreras de Informática (PROMINF, 2013-2017), donde uno de los ejes de trabajo fue el mejoramiento de indicadores de ingreso, permanencia y promoción para la carrera de Ingeniería de Sistemas (EXA). Para ello, se puso en marcha un programa con tutores pares, que luego fue evolucionando y se relevaron los resultados para construir esos indicadores que marcaron el avance y resultado del proyecto.

-El Mag. Claudio Aciti es el actual Vicedecano y secretario general de EXA (desde 2017) donde una de las áreas que coordina es la CAEG que se encarga del seguimiento del ingreso, permanencia y egreso.

-El Dr. Guillermo Rodríguez ha participado en numerosos artículos relacionados a las técnicas de *Educational Data Mining*.

-La Médica Veterinaria María Pía Silvestrini es co-directora del proyecto de investigación "Desarrollo de un sistema de indicadores para el monitoreo del desempeño académico en las carreras de Lic. en Enfermería y Medicina. ESCS-UNCPBA".

3. RESULTADOS OBTENIDOS/ESPERADOS

A pesar de que la investigación se desarrolla en una UA en particular, la investigación constituirá un experimento trasladable al resto. Por otra parte, se espera que los resultados de las investigaciones colaboren con las problemáticas referidas a la Responsabilidad Social Universitaria (RSU) aportando indicadores como consecuencia de la aplicación de las técnicas mencionadas precedentemente.

Se listan a continuación algunos de los trabajos en desarrollo especificando la técnica utilizada:

- Análisis del rendimiento académico en la asignatura Investigación Operativa. Clasificador de Bayes.

- Deserción académica (abandono) en la carrera de Ing. de Sistemas. Análisis multivariado.
- Deserción académica (abandono) en la carrera de Ing. de Sistemas. Reglas de asociación.
- Análisis de recusantes en la asignatura Investigación Operativa. Estadística básica (dispersión, correlación, regresión lineal) y Análisis multivariado.
- Análisis de permanencia en la carrera de Ing. de Sistemas. *Educational Data Mining*.
- Análisis del desgranamiento universitario en las cohortes estudiantiles de Ing. de Sistemas. Cadenas de Markov de orden uno con técnicas de validación.
- Desarrollo de un tablero de control con Indicadores de deserción. Inclusión de métodos de pronóstico.
- Estudio de la retención/deserción en Ing. de Sistemas. Técnicas de análisis causal.

Para el caso puntual de los trabajos donde se aplican datos sobre la asignatura de grado Investigación Operativa, se pretende analizar los resultados de estos estudios preliminares a fin de evaluar si es posible generalizarlo hacia otras asignaturas de la carrera Ing. de Sistemas.

4. FORMACIÓN DE RECURSOS HUMANOS

La estructura del equipo de trabajo se muestra en la siguiente tabla:

Apellido y nombre	Título	Cargo	Funciones
Aciti, Claudio	Mg.	Prof. UNCPBA	Integrante
Illescas, Gustavo	Dr.	Prof. UNCPBA	Director
Rodríguez, Guillermo	Dr.	Ayud. UNCPBA	Integrante
Silvestrini, Pía	Med. Vet.	Prof. UNCPBA	Integrante
Todorovich, Elías	Dr.	Prof. UNCPBA	Integrante

-El Mg. Claudio Aciti es doctorando en el Doctorado en Matemática Computacional e Industrial (DCMI) de EXA bajo la dirección del Dr. Todorovich E.

-La Médica Veterinaria María Pía Silvestrini (perteneciente al IICS) ha presentado su plan de doctorado en “Interacción entre aprendizaje colaborativo, aprendizaje basado en proyectos y educación centrada en la comunidad como estrategia para la formación por competencias del estudiante de medicina”. Directores: Marín G - Illescas G.

Becarios

- Indicadores de vinculación social en la carrera de Medicina UNICEN (2021-2022) Beca Estímulo a las Vocaciones Científicas (Ev-CIN). Dirección: Silvestrini M. P. Becario: Juan Manuel Ciancio Pagano.
- Análisis de la trayectoria educativa en Salud Pública mediante minería de datos (*Educational Data Mining*). Beca Iniciación a la investigación (Fortalecimiento UNCPBA 2021-2022) Dirección: Silvestrini M. P. Becario: Gimena Dietrich. Título.
- Análisis de Redes Sociales (ARS) para la mejora de las prácticas socioeducativas. Beca Ev-CIN (presentada 2022). Dirección: Silvestrini M. P. Becario: Juan Emilio Ciancio Pagano.
- Virtual Moodle Bot (VMB): Un Bot virtual en Moodle para asistir y complementar actividades académicas en espacios de educación superior. (presentada 2022). Dirección: Illescas G., Rodríguez G. Becario: Herrera De Rosa Exequiel.

Tesis de grado

- Herramienta web para el análisis y descubrimiento de indicadores. Dirección: Illescas G. Tesista: Juan Disteffano.
- Análisis del rendimiento académico en la carrera de Ing. de Sistemas utilizando clasificadores (Random forest, K-nn vecinos más cercanos, Support vector machine y Bayes). Dirección: Illescas G., Mora-Soto A (Univ. Celaya MEX.). Tesistas: Maylen Dell Oso, Florencia Paglione.

5. BIBLIOGRAFÍA

- Bañuelos Saucedo, A. L. (2013). Análisis del avance escolar a nivel licenciatura aplicando cadenas de Markov. Universidad Nacional

- Autónoma de México (UNAM). Tesis Maestría.
- Brown, Tim (2009). *Change by Design, How Design Thinking Transforms Organizations and Inspires Innovation* Ed. Harper Collins New York ISBN 978-0-06-193774-3
 - Carlos, H.; German, S.; Dixon, S. (2021). *Tashi-Bot: A Intelligent Personal Assistant for Users in an Educational Institution*. 10.20944/preprints202108.0380.v1.
 - Chinedu O., Ade-Ibijola A. (2021). *Chatbots applications in education: A systematic review*. *Computers and Education: Artificial Intelligence*. 2. 100033. 10.1016/j.caeai.2021.100033.
 - Concessao R. (2016) *Predictive Data Analytics: for informes strategic decisions*. Editor: CreateSpace Independent Publishing Platform. Edición: 5.0. 2016.
 - Eugenia, Á. R. (2000). *Modelo markoviano para el estudio de evolución de cohortes de estudiantes de un programa académico*. Red de revistas científicas de América Latina y Caribe, España y Portugal.
 - Gutiérrez Pizarro C. (2014). *Estudio de la permanencia y deserción de los estudiantes de pregrado en una Facultad de Ingeniería, utilizando una metodología mixta de Investigación y árboles de decisión en su predicción*. Tesis doctoral Univ. de Alcalá.
 - Hernández García C., Rodríguez Rodríguez J. (2016). *Algoritmo híbrido basado en aprendizaje computacional para el manejo de datos faltantes en aplicaciones OLAP Ingeniare*. *Revista chilena de Ing.*
 - Hernández Leal P. F. (2011). *Algoritmo de Aprendizaje para Redes Bayesianas de Nodos Temporales*. Instituto Nacional de Astrofísica, Óptica y Electrónica. Ed. INAPE. México.
 - Miller T. (2015), *Marketing Data Science: Modeling Techniques in Predictive Analytics with R and Python*, Editorial Pearson Education, versión impresa y Kindle.
 - Miranda M y Guzmán J. (2017). *Análisis de la Deserción de Estudiantes Universitarios usando Técnicas de Minería de Datos-Form*. Univ. vol.10 no.3 La Serena. Chile
 - Morrison D. (2007). *A Naive Bayes: Decision Tree Hybrid*
 - Oviedo B., Moreno A., Puris A., Villacís A., Delgado D. (2015). *Análisis de datos educativos utilizando Redes Bayesianas*. 13th Latin American and Caribbean Conference for Engineering and Technology. Santo Domingo, República Dominicana.
 - Papoulis A., Pillai U. (2002). *Probability Random Variables and Stochastic Processes*. McGraw-Hill.
 - Saaty, T. 2008. *Decision Making with the Analytic Hierarchy Process*, Int. J. Services Sciences, Vol. 1, Nº. 1, pp.83-98
 - Siegel E. (2016), *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die*. Ed. Wiley.
 - SIU (Sistema de Información Universitaria) (2021) <https://portal.comunidad.siu.edu.ar/>
 - Vegega C., Deroche A., Pytel P., Ramón H., Straccia L., Acosta M., Pollo-Cattaneo M (2018). *Diagnóstico del Proceso de Aprendizaje de Alumnos de Inteligencia Artificial mediante un Modelo Dinámico Bayesiano*. *Rev. Tecnología y Ciencia*. UTN.

CONSTRUCCIÓN DE CONOCIMIENTOS EN LA UNIVERSIDAD A PARTIR DEL DISEÑO E IMPLEMENTACIÓN DE ESCENARIOS TECNO-PEDAGÓGICOS Y MATERIALES EDUCATIVOS DIGITALES

Pontoriero Francisco: ruffopontoriero@gmail.com;

Ureta Laura: laubureta@gmail.com;

Cuadros Patricia: pcuadrosan@hotmail.com

Grupo de investigación perteneciente al Gabinete de Tecnología e Innovación Educativa.

Dpto. de Informática - Dpto. de Geofísica y Astronomía. - Dpto. de Geología - Dpto. de Biología de la Facultad de Ciencias Exactas, Físicas y Naturales y Dpto. de Matemáticas de la Facultad de Ingeniería - Universidad Nacional de San Juan.

RESUMEN

Los jóvenes tienen el mundo en las yemas de los dedos a través de las tecnologías; ellas atraviesan sus modos de pensar, aprender y conocer, razón por la cual es de relevancia que las instituciones educativas, docentes y responsables de su formación identifiquemos y reconozcamos sus tendencias culturales. Ante esta realidad, cabe preguntarnos, ¿Cuáles son las preferencias de nuestros alumnos en relación al uso de las tecnologías? ¿Qué materiales educativos digitales eligen? ¿Se inclinan por lo hipertextual, por el texto plano, por lo multimedial? ¿Prefieren los grandes relatos o las micro narrativas? ¿La intencionalidad pedagógica de nuestras prácticas de enseñanza atiende a funciones cognitivas como resolución de problemas, pensamiento estratégico, hipotetización, análisis de diferentes contextos, entre otras?

El eje de la investigación gira en torno a la generación de conocimientos acerca de qué residuos cognitivos dejan en alumnos universitarios el uso de macro/micro narrativas que favorezcan la construcción de conocimientos. A partir de estos conocimientos se apunta a generar pautas que orienten la evaluación de materiales educativos digitales.

Se empleó la metodología cualitativa, con una estrategia metodológica que combina el uso de técnicas de recolección de información como la investigación documental, la entrevista semiestructurada y otras. Se prevé utilizar fuentes de información primarias y secundarias.

Palabras Clave: Educación y Tecnología; Narrativas; Universidad; Materiales Educativos.

CONTEXTO

El presente Proyecto - aprobado por CICITCA - tiene como objetivo principal generar conocimientos que aporten a la problemática: qué efectos cognitivos dejan en alumnos universitarios los materiales educativos digitales.

Se encuentra enmarcado en el “Gabinete de Tecnología e Innovación Educativa” creado por Resolución 03/18, dependiente de Decanato de la Facultad de Ciencias Exactas, Físicas y Naturales. El Gabinete se constituye como un espacio académico y de investigación cuyo objetivo es generar estrategias, metodologías, medios y materiales, así como instrumentos tecnológicos, que contribuyan a la inserción de innovaciones tecnológicas en pos de la mejora de los procesos educativos de la Facultad de Ciencias Exactas, Físicas y Naturales y de la Facultad de Ingeniería de la Universidad Nacional de San Juan.

1. INTRODUCCIÓN

No es novedad que las tecnologías de la información y comunicación han invadido de diferentes maneras y con distintos formatos nuestra vida cotidiana y especialmente la educación. Los jóvenes tienen el mundo en las yemas de los dedos a través de las tecnologías; ellas atraviesan sus modos de pensar, aprender y conocer, razón por la cual es de relevancia que las instituciones educativas, docentes y responsables de su

formación identifiquemos y reconozcamos sus tendencias culturales. La inmediatez, lo instantáneo, lo nuevo, las diversidades predominan en sus actividades. Aquí es pertinente considerar la propuesta de Maggio (2012), quién expresa “*lo primero que debemos hacer es reconocer lo que los alumnos hacen*” (p.153), preguntarnos qué son capaces de hacer, cuáles son sus preferencias tecnológicas, en qué actividades las usan y cómo usan las tecnologías, qué valor les asignan en sus procesos de estudio y producción. Esto significa reconocer el valor de los usos culturales y formas cognitivas del mismo modo como deberíamos tener en cuenta los estilos cognitivos propios de los sujetos culturales que son nuestros alumnos.

Los resultados significativos en el aprendizaje pasan por dar prioridad a los aspectos didáctico-pedagógicos y no a los aspectos tecnológicos únicamente. La claridad de las metas pedagógicas acompañada de una metodología de aprendizaje pertinente son factores clave para la selección de las tecnologías que contribuirán con la comprensión de las temáticas abordadas. En el nuevo contexto socio-tecnológico, las diversas formas de construcción de conocimiento con tecnologías digitales requieren de la integración de habilidades a fin de que el estudiante pueda buscar, encontrar, analizar, evaluar y utilizar eficazmente la información que necesita, pero

además transformar, comunicar y colaborar con sus pares en la resolución de problemas. A partir de esto, resulta evidente la necesidad de empoderar a nuestros alumnos con capacidades que los ayuden a mejorar su autonomía, la toma de decisiones, la identificación de sus propias necesidades. Ante esta realidad, cabe preguntarnos, ¿Cuáles son las preferencias de nuestros alumnos en relación al uso de las tecnologías? ¿Usan y participan activamente en las redes sociales? ¿Qué materiales educativos digitales eligen? ¿Se inclinan por lo hipertextual, por el texto plano, por lo multimedial? ¿Prefieren los grandes relatos o las micro narrativas? ¿La intencionalidad pedagógica de nuestras prácticas de enseñanza atiende a funciones cognitivas como resolución de problemas, pensamiento estratégico, hipotetización, análisis de diferentes contextos, entre otras? ¿Al utilizar tecnologías de diferente complejidad, priorizamos el efecto de la tecnología? Los docentes tenemos un desafío importante; es necesario que apostemos al uso y/o generación de materiales educativos digitales que propongan estrategias y actividades educativas abiertas y flexibles que favorezcan la construcción de conocimiento. Somos conscientes de la necesidad de formar recursos humanos creativos, que sean capaces de anticiparse a los cambios y a las problemáticas, que aprendan a saborear la búsqueda de conocimiento y la inquietud de aplicarlo a la acción.

En base a lo expuesto, el proyecto de investigación tiene como objetivos específicos:

- Identificar características de macro y micro narrativas como materiales educativos digitales en carreras universitarias.
- Diseñar materiales educativos digitales y e-actividades que favorezcan la construcción de conocimiento para diferentes áreas disciplinares.
- Identificar el residuo cognitivo en los procesos de aprendizaje de alumnos a partir del uso de materiales educativos digitales: Macro y Micro narrativas.
- Generar pautas que orienten la evaluación de materiales educativos digitales.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Se busca generar conocimiento en relación con los efectos cognitivos que dejan los materiales educativos digitales en estudiantes universitarios

3. RESULTADOS OBTENIDOS/ESPERADOS

Se lograron identificar características de macro y micro narrativas en vistas al diseño de materiales educativos. Se crearon escenarios tecno-pedagógicos con el objeto de favorecer la construcción de conocimiento, dentro de los cuáles se diseñaron diferentes materiales educativos digitales.

Los materiales educativos digitales (tales como videos, archivos de texto plano, objetos de aprendizaje con exelearning y e-actividades como foros) se elaboraron con distintos formatos, teniendo en cuenta la intencionalidad pedagógica y el área de conocimiento. En el marco de una investigación de “diseño”, tal variedad de materiales nos permitió construir narrativas que orienten a nuestros alumnos a desarrollar conocimiento como sujetos culturales en un paradigma socio-tecnológico.

Además, se aplicaron encuestas para recolectar información sobre los resultados obtenidos por los alumnos en sus procesos de aprendizaje, en el contexto de estos escenarios tecno-pedagógicos.

Las siguientes actividades se centrarán en el procesamiento de la información recolectada, en pos de identificar el residuo cognitivo generado en los alumnos como resultado de los procesos de aprendizaje a partir del uso de materiales educativos digitales. Esto permitirá, además, generar pautas que orienten la evaluación de materiales educativos digitales.

El análisis mencionado se llevará a cabo a la luz de la bibliografía vinculada con el presente tema de investigación.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está conformado de la siguiente manera:

- Director: Pontoriero, Francisco
- Co-directora: Ureta, Laura
- Investigadores: González, Liliana Mirna, Cuadros Patricia, Margarit Viviana, Gallardo Vanesa, Marcovecchio, María J., Rossetti Beiram, Gabriela.

Se propone volcar los resultados en actividades de Postgrado, tales como:

- Diseño e implementación de la Diplomatura “Educación y nuevas tecnologías en tiempos de convergencia tecnológica”. Aprobada por Ordenanza N° 5 CD-FCEFN, Prof. Responsable: Mgter Liliana Mirna González.
- Co-dirección/Asesoramiento de Tesis de Maestría en Informática. F.C.E.FyN de la U.N.S.J.
- Docente de la Maestría en Informática. F.C.E.FyN de la UNSJ. Materia: Educación y Tecnologías.
- BECARIOS
Puigdengolas Márquez, Fabricio Agustín
Facultad de Ingeniería. Universidad Nacional de San Juan. Tareas de investigación y desarrollo. Co-director o co-tutor: Cuadros, Patricia Del Valle.

En paralelo, se buscará desarrollar instancias de interacción con las cátedras interesadas de carreras de la F.C.E.F y N. de la U.N.S.J.,

para orientar en el diseño de materiales educativos, junto con la ejecución de talleres de análisis y discusión de la problemática bajo estudio.

5. BIBLIOGRAFIA

- Ambrosino, María Alejandra (2017). *Docencia y NARRATIVAS Transmedia en la educación superior*. Trayectorias Universitarias. Volumen 3, N° 4. Recuperado de <http://revistas.unlp.edu.ar/TrayectoriasUniversitarias>
- Canales Cruz Alejandro (2014). *Hacia un nuevo diseño para el aprendizaje: escenarios educativos para la Web 2.0*. Revista de Innovación Educativa. Vol. 6, Núm. 2
- Maggio M. (2012) *Enriquecer la Enseñanza*. Paidós. Argentina.
- Maggio M. (2018) *Reinventar la clase en la Universidad*. Paidós. Argentina
- Odetti, V. (2016). *Materiales didácticos hipermediales: lecciones aprendidas y desafíos pendientes*. En García, J. M., y Báez Sus, M. (Comp.) Educación y tecnologías en perspectiva. Diez años de FLACSO Uruguay. Disponible en http://flacso.edu.uy/publicaciones/libro_educacion_tecnologia_2016/Baez_Garcia_Educacion_y_tecnologias_en_perspectiva.pdf
- Odetti, V. (2017). *El diseño de materiales didácticos hipermediales: El caso del PENT FLACSO*. Buenos Aires: Teseo.
- Odetti, V. (2018). *Narrativas Transmedia. Educación y TIC*. EL ABROJO. Montevideo. Uruguay
- Oliva Elisa; Gallardo Vanesa; Ciancio María Inés; Salas Andrea (2021). *Identificación de Regularidades en Matemática, mediante determinación de patrones y con uso de software libre*. Workshop. XXIII Edición del Workshop de Investigadores en Ciencias de la Computación. La Rioja (Chilecito).
- Pontoriero Francisco (2020). *La Educación Superior a Distancia en el Marco Legal Argentino*. Psico/pedagógica, psicología y pedagogía de la persona. Mendoza: Centro de Investigaciones Cuyo, 2020. vol. 14, ISSN 0328-5413
- Pontoriero Francisco, González Liliana, Ureta Laura, Margarit Viviana, Marcovecchio María José (2020). *Materiales Educativos Digitales: Macro Narrativas y Micro Narrativas para la Construcción de Conocimientos en la Universidad*. XXII Workshop de Investigadores en Ciencias de la Computación - WICC 2020. El Calafate.. Universidad Nacional de la Patagonia Austral. <http://www.wicc2020.unpa.edu.ar>
- Scolari, C. (2014). *Narrativas transmedia: nuevas formas de comunicar en la era digital*. Anuario AC/E de cultura digital. Recuperado de <https://www.accioncultural.es/media/Default%20Fi>

les/activ/2014/Adj/Anuario_ACE_2014/6Tr
ansmedia_CScolari.pdf

Modelos de desarrollo de serious games Las analíticas de aprendizaje e inteligencia artificial

Stella Maris Massa¹, Lucrecia Ethel Moro, Gustavo Bacino, Guccione Leonel, Hernán Hinojal, Adolfo Spinelli, Esteban Zapirain, Carlos Rico, Franco David Kühn y Franco Lanzillota.

¹Facultad de Ingeniería/ Universidad Nacional de Mar del Plata/Argentina
(7600) Av. Juan B. Justo 4302, +54-223-4816600
smassa4@gmail.com, lucreciamoro@gmail.com, gustavobacino@gmail.com,
leonel.guccione@gmail.com, hinojal@fi.mdp.edu.ar, spinelliadolfo@gmail.com,
estebanzapirain@gmail.com, carlos@fi.mdp.edu.ar, fdkuhn@gmail.com,
franco.lanzi96@gmail.com

RESUMEN

En este trabajo se presenta el proyecto de investigación, continuación de uno anterior denominado: “El Proceso de Desarrollo de Serious Games. Modelos, herramientas y Analíticas de Aprendizaje”, en el cual se amplía el alcance de la interacción entre el jugador (estudiante) y el Serious Game (SG), aportando una retroalimentación basada en las métricas que arrojan las Analíticas de Aprendizaje (AA) del desempeño logrado durante el desarrollo del juego. Con la posibilidad de procesar grandes volúmenes de datos, los sistemas de recomendación, son un enfoque práctico para proporcionar la información más adecuada al usuario en función de su comportamiento y contexto.

Este proyecto tiene como objetivo construir un marco teórico-práctico para la incorporación de recomendaciones en un SG a partir de las métricas de evaluación obtenidas de las AA.

Nuestra propuesta en este proyecto, se enfoca en adaptar y refinar el SG utilizando técnicas de Inteligencia Artificial y mejorando el análisis de los procesos relacionados con el aprendizaje.

Se combinan tres tendencias principales en la investigación del aprendizaje con tecnologías: Serious Games, Analíticas de Aprendizaje e Inteligencia Artificial (IA).

Palabras clave: Modelos de desarrollo, Analíticas de Aprendizaje, Inteligencia Artificial, Serious Games.

CONTEXTO

En el presente proyecto, continuación del anterior: “El Proceso de desarrollo de Serious Games. Modelos, Herramientas y Analíticas de Aprendizaje” (15/G585), se profundiza el proceso de análisis de los resultados arrojados mediante AA, en un SG, para la obtención de métricas relacionadas con el desempeño (estudiante /aprendiz y docente/instructor), en los diversos contextos de aplicación como defensa, atención médica, educación, gestión de emergencias, planificación urbana e ingeniería.

En el mismo, se amplía el alcance de la interacción entre el jugador y el SG aportando una retroalimentación al docente (instructor) y al estudiante (aprendiz) basada en las métricas de la performance o desempeño logrado en el desarrollo del juego.

INTRODUCCIÓN

El aporte de los videojuegos al aprendizaje se ve evidenciado en muchos aspectos. Los jugadores aprenden a obtener información de múltiples fuentes, a tomar decisiones rápidamente, a deducir las reglas del juego, a crear estrategias para superar obstáculos, a experimentar y comprender sistemas complejos y a trabajar colaborativamente [1].

Estas ventajas hacen de los SG una gran oportunidad para enseñar distintas habilidades y situaciones del mundo real. El objetivo final de un SG es la diversión, inmersión y aprendizaje. Si bien el propósito de todo proceso conducente a formar o entrenar en algún aspecto es el aprendizaje, la diversión, como sinónimo de esparcimiento, es un aspecto que debe estar presente en un SG precisamente porque su propio nombre lleva incorporada la palabra juego (en lo que parece un oxímoron, por aquello de juego serio). Si bien su objetivo principal no es el entretenimiento, esto no significa que los SG no puedan contener esa característica [2].

Un caso particular de SG son los Simuladores de Entrenamiento (TS), los cuales son especialmente utilizados en aquellos dominios donde se requiere que las personas manipulen o controlen sistemas complejos [3]. Estos simuladores están siendo ampliamente utilizados en diversos ámbitos, como el adiestramiento militar, el entrenamiento civil y las aplicaciones comerciales [4].

Actualmente, las AA resultan de mucho interés a la hora de obtener resultados cuantitativos respecto del desempeño de los estudiantes o aprendices. Si bien su inclusión dentro de los SG se encuentra en constante crecimiento, aún se ha trabajado poco sobre su uso puntual en los TS.

Las AA tienen su origen en los conceptos de inteligencia empresarial, el uso de datos en una empresa para facilitar la toma de decisiones. Estas ideas han sido apropiadas por los ámbitos educativos con el objetivo de brindar nuevas herramientas que permitan comprender y optimizar el aprendizaje, así como los entornos en los que este tiene lugar [5].

Hershkovitz et al. [6] señalan que las AA han tenido un enorme progreso en los últimos años, debido a que se han ido incorporando métodos estadísticos avanzados y técnicas de aprendizaje automático. Esto ha permitido que se puedan capturar y analizar una mayor cantidad de datos, así como también ha

mejorado la calidad de la información que se obtiene como resultado de dichos análisis.

En proyectos anteriores [7 - 9] hemos investigado la integración de las AA en los SG y desarrollado la elicitación de los objetivos pedagógicos y los elementos básicos de jugabilidad e inmersión de un SG [10]. Se han definido los requerimientos del mismo y las habilidades a ser desarrolladas por los estudiantes, conjuntamente con las variables y eventos relacionados con ellas [11]. Finalmente, se ha concluido en el desarrollo de SG a los cuales se le han incorporado AA para poder obtener información relevante del proceso de aprendizaje de los estudiantes [12].

Se trabajó con la metodología de Kitchenham [13] para el desarrollo de una revisión sistemática de la literatura en relación a la integración de AA en TS. A partir de los artículos hallados, se ha comprobado que, si bien la temática se encuentra en crecimiento, aún no existe mucha información disponible y son pocos los desarrollos que se han realizado [14].

En el caso de los TS, la evaluación del aprendizaje y de las habilidades adquiridas durante la simulación es una tarea muy compleja. Por lo tanto, hacen falta métodos que permitan recopilar y analizar los datos resultantes de las interacciones de los aprendices, para poder así medir su desempeño y producir evidencias válidas del aprendizaje.

En este marco, se propuso llevar a cabo la puesta en valor de un TS para el uso de artillería antiaérea, incorporando mejoras en la simulación con el fin de aumentar la experiencia inmersiva, conjuntamente con la medición de datos y presentación de informes. Este sistema inspecciona cómo interactúa cada aprendiz con el simulador, almacenando información detallada y generando informes sobre las interacciones y los cambios en el estado interno de la simulación para un análisis posterior, a partir de la integración de las AA en el TS. De esta manera, permite al

instructor evaluar objetivamente el desempeño del aprendiz, además de poder medir su progreso a lo largo de las sesiones de entrenamiento [15]. Como casos de estudio, podemos mencionar los trabajos realizados y publicados por integrantes de este Grupo de investigación (Grupo de investigación en tecnologías interactivas, GTI): [16 – 17].

Estas características están alineadas con las tendencias claves, los desafíos y los desarrollos importantes en tecnología educativa propuestos en el informe anual Horizon (2021) para educación superior [18]. Dicho informe incluye "seis tecnologías pronosticadas como importantes para la enseñanza, el aprendizaje y la investigación creativa en el futuro". A corto plazo estas son: inteligencia artificial (IA), modelos de cursos mixtos e híbridos, analítica del aprendizaje, micro-credenciales, recursos educativos abiertos (REA) y calidad en el aprendizaje en línea.

Ya el informe anual Horizon 2012, señalaba a las AA como uno de los campos a tener en cuenta a mediano plazo, definiéndolas como "el análisis de una amplia variedad de datos producidos por estudiantes, con el objetivo de evaluar su progreso académico, predecir sus futuros resultados e identificar riesgos potenciales" [19].

Riedel, Essa & Bowen [20], en su publicación de EDUCAUSE definieron la IA como "sistemas informáticos que realizan tareas que generalmente se cree que requieren procesos cognitivos humanos y capacidades de toma de decisiones", y esa definición todavía parece adecuada.

Si bien la pandemia del Covid-19 ha detenido prácticamente a industrias enteras, la IA parece haber podido mantener su progreso a un ritmo casi vertiginoso, y el peligro de otro "invierno de la IA" parece remoto. Según el Índice de IA de 2021, el progreso en el procesamiento del lenguaje natural ha sido tan rápido que los avances técnicos han comenzado a superar los puntos de referencia

para probarlos. La IA ha avanzado hacia el "aprendizaje autosupervisado".

Tal como indica Alonso-Fernández [21], la manera particular en que se obtengan los datos determinará profundamente las posibilidades tanto en el análisis como en la visualización, ya que implicará la cantidad de información que se recibirá y la riqueza de la misma.

Las visualizaciones de resultados también deben ajustarse a las necesidades de los diferentes agentes involucrados, estudiantes y docentes, como parte del uso educativo del juego, así como para desarrolladores o gestores, que estarían más interesados en el correcto funcionamiento del juego, pero también desean conocer si éste está cumpliendo con los objetivos de aprendizaje previstos.

Con la posibilidad de procesar grandes volúmenes de datos (big data, minería de datos educativos, AA), los sistemas de recomendación son un enfoque práctico para proporcionar la información más adecuada al usuario en función de su comportamiento y contexto.

Duval [22] introdujo las recomendaciones como una solución "para lidiar con la paradoja de la elección". La retroalimentación tiene el potencial de apoyar el rendimiento académico, promover la motivación, el aprendizaje auto-supervisado y la autoeficacia de los estudiantes, con la finalidad de acortar la brecha entre su desempeño actual y el deseado por ellos [23].

Tal como señalan Westera, Nadolski y Hummel [24], el análisis de los datos del desempeño de los estudiantes/aprendices en los SG posee dos diferentes etapas. En la primera, se analiza la información recopilada de cada usuario individualmente durante el juego (mediante las AA), con el fin de mejorar y personalizar su interacción con el sistema. En la segunda etapa, posterior al juego, se trabaja con la información obtenida de las AA para adaptar el SG generando nuevas intervenciones que contribuyan a

mejorar el aprendizaje. Nuestra propuesta en este proyecto, se enfoca en adaptar y refinar el SG utilizando técnicas de IA, mejorando el análisis de los procesos relacionados con el aprendizaje en el SG.

En ese marco, se establecen los siguientes objetivos específicos:

- Consolidar y estandarizar las actividades que conforman el proceso de incorporación de Analíticas de Aprendizaje en los Serious Games.
- Sistematizar el proceso de generación de reportes que visualicen la información relevante del proceso de aprendizaje de los estudiantes sobre la base de sus datos de interacción en Serious Games.
- Diseñar e implementar un sistema adaptativo en los Serious Games incorporando recomendaciones mediante técnicas de Inteligencia Artificial a partir de las Analíticas de Aprendizaje.
- Desarrollar un trabajo de campo en el que se analizará la viabilidad y posibilidades del Proceso de integración de recomendaciones en Serious Games.

La mayoría de las publicaciones en el ámbito de la línea de investigación que abordamos están enfocadas en la implementación de estas tecnologías en el ámbito educativo, pero las soluciones propuestas son varias, sin encontrarse lineamientos generales más allá de la definición de las propias AA.

Sistematizar los procesos de producción de SG, particularmente la selección de estrategias para extraer información relevante del proceso de aprendizaje del jugador/estudiante, contribuirá con la difusión de buenas prácticas en un sector en expansión y permitirá la apertura de líneas de investigación mediante la elaboración de un marco de referencia sobre metodologías y tecnologías emergentes.

FORMACIÓN DE RECURSOS HUMANOS

La formación de recursos humanos se considera un aspecto imprescindible e insoslayable. Integrantes del proyecto se encuentran desarrollando y dirigiendo seis tesis de postgrado, en el marco del proyecto de investigación, correspondientes al Doctorado en Humanidades y Artes- Mención Ciencias de la Educación (UNR); al Doctorado en Modelado y Simulación Computacional (UNMdP); a la Maestría en Ingeniería de Software (UNLP) y a la Especialización en Docencia Universitaria (UNMdP).

REFERENCIAS BIBLIOGRÁFICAS

- [1] Prensky, M. (2003). Digital game-based learning. *Computers in Entertainment (CIE)*, 1(1), pp. 21–21.
- [2] Massa, S.M. (2017). Serious Games y aprendizaje. Nuevos desafíos educativos en Massa, S.M. y Bacino, G.A. (comp.), *Videjuegos en Serio. Creando Serious Games para aprender jugando*, pp. 11-26. Mar del Plata, Argentina: Universidad Nacional de Mar del Plata.
- [3] Drews, F. A. & Bakdash, J. Z. (2013). Simulation Training in Health Care. *Reviews of Human Factors and Ergonomics*, 8(1), pp.191–234.
- [4] Camarasa, D. A. y Bianchi, O. M. (2012). Desarrollo de Software de Simulación Inmersiva para Fracciones Heterogéneas. XIV Workshop de Investigadores en Ciencias de la Computación (WICC), Posadas, Misiones, Argentina.
- [5] Ferguson, R. & Shum, S. B. (2012). Social learning analytics: Five Approaches, LAK '12: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, 2012, pp. 23–33.
- [6] Hershkovitz, A., Knight, S., Dawson, S., Jovanovic, J. & Gasevic, D. (2016). About 'learning' and 'analytics'. *Journal of Learning Analytics*, 2(3), pp. 1–5.
- [7] Massa, S.M. y Kühn, F. (2018). Analíticas de Aprendizaje para Serious Games. XX Workshop de Investigadores en Ciencias de la Computación (WICC), Corrientes, Corrientes, Argentina, 2018.
- [8] Kühn, F. y Massa, S.M. (2018). Learning Analytics in Serious Games: a systematic review of literature. IEEE ARGENCON 2018 Congress, San Miguel de Tucumán, Tucumán, Argentina.

- [9] Massa, S. M., Moro, L., Bacino, G., Pirro, A., Evans, F., Hinojal, H., Spinelli, A., Zapirain, E., Rico, C., Kühn, F. y Lanzillotta, F. (2020). El proceso de desarrollo de Serious Games Modelos, herramientas y analíticas de aprendizaje. WICC 2020, XXII Workshop de Investigadores en Ciencias de la Computación. El Calafate, Santa Cruz, Argentina. Universidad Nacional de la Patagonia Austral. RedUnci. Publicado en Actas, 878-882, 2020. ISBN 978-987-3714-82-5.
- [10] Spinelli A. y Massa, S. M. (2018). Elicitation in Serious Game. IEEE ARGENCON 2018 Congress, San Miguel de Tucumán, Tucumán, Argentina.
- [11] Spinelli, A., Massa, S.M., Rico ,C. y Kühn F. (2018). Diseño de Serious Games. Requerimientos del Juego – Competencias y Habilidades. XX Encuentro Internacional Virtual Educa, Buenos Aires, Argentina, 2018.
- [12] Kühn, F. (2019). Juegos Serios y Analíticas de Aprendizaje: Implementación en el entorno educativo. Trabajo Final de Grado, Facultad de Ingeniería, UNMDP, Mar del Plata, Buenos Aires, Argentina.
- [13] Kitchenham, B. (2004). Procedures for Performing Systematic Reviews. Joint Technical Report. Keele University TR/SE-0401 and NICTA 0400011T.1.
- [14] Lanzillotta, F., Massa, S.M. y Spinelli, A. (2020). Learning analytics in training simulators: a systematic review of literature. IEEE ARGENCON 2020. Resistencia, Chaco, Argentina, Publicado en Actas, 1-872, 2020. DOI:10.1109/ARGENCON49523.2020.9505456.
- [15] Lanzillotta, F. (2020). Puesta en valor de un simulador de entrenamiento mediante la incorporación de experiencia inmersiva y analíticas de aprendizaje. Trabajo Final de Grado, Facultad de Ingeniería, UNMDP, Mar del Plata, Buenos Aires, Argentina.
- [16] Fantini, L.M., Spinelli, A.T. y Kühn, F. (2021). Juegos serios en el proceso de aprendizaje de lectoescritura. Encuentro Argentino y Latinoamericano de Ingeniería. 5° Congreso Argentino de Ingeniería. (CADI). 11° Congreso Argentino de Enseñanza de la Ingeniería (CAEDI). 3° Congreso Latinoamericano de Ingeniería (CLADI). Buenos Aires, Argentina.
- [17] Zapirain, E., Kühn, F., Agüero, M., Pieroni, A. y Echeverría Luchini, L. (2021). Diseño de un juego serio como herramienta de revitalización cultural. Encuentro Argentino y Latinoamericano de Ingeniería. 5° Congreso Argentino de Ingeniería. (CADI). 11° Congreso Argentino de Enseñanza de la Ingeniería (CAEDI). 3° Congreso Latinoamericano de Ingeniería (CLADI). Buenos Aires, Argentina.
- [18] Pelletier, K., Brown, M., Brooks, D., McCormack, M., Reeves, J., Arbino, N., Bozkurt, A., Crawford, S., Czerniewicz, L., Gibson, R., Linder, L., Mason, J. & Mondelli, V. (2021). EDUCAUSE Horizon Report 2021. Teaching and Learning Edition (Boulder, CO: EDUCAUSE, 2021).
- [19] Johnson, L., Adams Becker, S. & Cummins, M. (2012). NMC Horizon Report: 2012 K-12 Edition. Austin, Texas: The New Media Consortium.
- [20] Riedel, K., Essa, A & Bowen, K. (2017). 7 Things You Should Know. About Artificial Intelligence in Teaching and Learning. EDUCAUSE Learning Initiative (ELI).
- [21] Alonso-Fernández, C. (2016). Gaming learning analytics for serious games. Trabajo fin de Grado. Facultad de Informática. Universidad Complutense de Madrid.
- [22] Duval, E. (2011). Attention please! Learning analytics for visualization and recommendation. Proceedings of the 1st International Conference on Learning Analytics and Knowledge (LAK '11), Banff, AB, Canada (pp. 9–17). New York: ACM.
- [23] Gajardo Muñoz, C. (2019). Falta de eficacia en la retroalimentación en el proceso evaluativo de los estudiantes. Tesis de grado Magister, Facultad de Educación, Universidad del Desarrollo, Chile.
- [24] Westera, W., Nadolski, R. y Hummel, H. (2014). Serious Gaming Analytics: What Students Log Files Tell Us about Gaming and Learning. International Journal of Serious Games, 1(2), pp. 35–50.

MODELO DE DISEÑO DE APRENDIZAJE PARA LA ENSEÑANZA DE LA FÍSICA EN UNA MODALIDAD E-LERNING USANDO LABORATORIOS VIRTUALES

Espeche Fabián, Durán Elena, Juárez Carlos

Instituto de Investigaciones en Informática y Sistemas de Información (IISI)

Facultad de Ciencias Exactas y Tecnologías (FCEyT)

Universidad Nacional de Santiago del Estero (UNSE)

e-mail: h_espeche@hotmail.com, eduran@unse.edu.ar, cjuarez@unse.edu.ar

CONTEXTO

En este trabajo se presenta una línea de investigación que se encuentra en desarrollo en el marco del proyecto de investigación “Modelos basados en Inteligencia Artificial y Computación Ubicua para la resolución de Problemas en Educación y otros dominios”, correspondiente a la convocatoria 2021 de la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Santiago del Estero (SICYT - UNSE). El proyecto tiene un período de ejecución desde el año 2022 hasta el año 2025 y busca favorecer el desarrollo de conocimiento científico- tecnológico de relevancia, principalmente sobre Computación Ubicua e Inteligencia Artificial, realizando propuestas de modelos computacionales para la resolución de problemas en Educación y otros dominios.

RESUMEN

La línea de investigación del presente trabajo surge luego de observar los problemas relacionados con la enseñanza de la Física en una modalidad e-learning, ya que la misma es una disciplina experimental. Esto nos pone en presencia de un problema educativo y tecnológico, que puede ser abordado desde la Informática Educativa. Desde este enfoque, esta línea de investigación se centra en el conocimiento pedagógico aplicado a los diseños de aprendizaje que realizan los docentes para el uso de los laboratorios virtuales para la enseñanza de la Física.

Se plantea comenzar con una investigación exploratoria en bibliotecas digitales con el fin de desarrollar el estado del arte de cómo se usan los laboratorios virtuales para la enseñanza de la Física en el nivel universitario y como se evalúa el uso de los mismos. Luego, con una investigación descriptiva se buscará caracterizar el uso que se le da a los laboratorios virtuales en ese ámbito, con el

fin de conocer los tipos de conocimiento pedagógicos utilizados, y la consideración de factores psicológicos, pedagógicos y tecnológicos en el diseño de las prácticas docentes. A partir de esa información, se definirán pautas para el diseño de aprendizajes en Laboratorios Virtuales de Física, que utilicen el conocimiento pedagógico y estimulen un aprendizaje meta-cognitivo, autónomo y autorregulado en los alumnos. Luego se planteará un modelo de diseño de aprendizajes, que cumpla con las pautas definidas, y se experimentará con este modelo en las prácticas docentes reales.

Palabras clave: Diseño de aprendizaje, Laboratorio Virtuales, Aprendizaje Metacognitivo, Aprendizaje Autorregulado. Modalidad e-learning.

1. INTRODUCCIÓN

Como consecuencia de la pandemia de COVID-19, que generó la suspensión de las actividades áulicas presenciales, quedaron al descubierto muchas problemáticas en el ámbito educativo en el nivel universitario. Se tuvo que pasar, de una educación presencial a una educación a distancia mediada por la Tecnología de la Información y la Comunicación (TIC). Fue así que muchas instituciones universitarias implementaron los Entornos Virtuales de Aprendizaje (EVA), como la manera de organizar y dar continuidad pedagógica a la trayectoria de los alumnos, transformando la modalidad de enseñanza presencial en una a distancia o e-learning.

Si nos centramos en la enseñanza de la Física en el nivel universitario, y considerando que esta es una ciencia en la que sus conceptos se basan en fenómenos físicos, que requieren un grado de abstracción para su enseñanza y aprendizaje; estos fenómenos pueden ser recreados o experimentados en los laboratorios institucionales, a través de las denominadas prácticas de laboratorios. Estas

prácticas permiten articular los conocimientos teóricos desarrollados en el aula con la experimentación que pueden realizar los alumnos en los laboratorios, algo que no es contemplado en una modalidad e-learning.

En la Web, podemos encontrar, los Laboratorios Virtuales (LV), que se definen como: “sistema informático que pretende simular el ambiente de un laboratorio real y que mediante simulaciones interactivas permite desarrollar las prácticas de laboratorio” (Cabrera Medina e Sánchez Medina, 2016). Estos venían siendo utilizados, antes de la pandemia de COVID 19, como recursos didácticos en apoyo a la enseñanza de distintos conceptos de la Física, pero siempre en una modalidad presencial, implementando distintas estrategias pedagógicas para las prácticas, de forma presencial o bimodal b-learning. En esta última se utiliza el encuentro presencial para la puesta en común del uso del LV (López et. al., 2016), (Arguedas-Matarrita et. al., 2017; Da Silva, 2012; Ardura y Zamora, 2013; Ré y Giubergia, 2013). Sin embargo, no se han encontrado evidencias del uso de laboratorios virtuales para la enseñanza de la Física en una modalidad e-learning.

Enfocando esta problemática desde el campo de la Informática Educativa (IE), es preciso hacer referencia al conocimiento pedagógico que debe estar presente en todo proceso de enseñanza y de aprendizaje, mediado por la tecnología. Cabe señalar que, IE se define como: “el desarrollo, uso y evaluación de sistemas digitales, que utilizan conocimiento pedagógico para participar en o facilitar el descubrimiento de recursos para soportar el aprendizaje” (Ford, 2008).

El conocimiento pedagógico incluye no sólo conocimiento para la enseñanza en un sentido tradicional, sino además el referido al “autoaprendizaje” o conocimiento “meta-cognitivo”, que las personas necesitan para participar en procesos de aprendizaje autónomos y autorregulados fuera de cualquier contexto institucional de educacional formal.

El conocimiento pedagógico utilizado por los sistemas informáticos educativos puede ser de distintos tipos y niveles de complejidad: metadatos específicos para educación, ontologías pedagógicas hasta estructuras de argumentación académica. Las técnicas utilizadas por estos sistemas, para procesar el conocimiento pedagógico, también varían desde complejos mecanismos de inferencia diseñados para descubrir recursos de aprendizaje, según las características y necesidades individuales

de los alumnos, hasta dispositivos diseñados para permitirles a los alumnos conocer y utilizar sus capacidades de meta-cognición.

El conocimiento pedagógico debe ser adaptado según la modalidad, ya que existen diferencias sustanciales entre la modalidad presencial y la modalidad a distancia o e-learning, generando problemáticas particulares; tales como, la interacción entre los actores, tiempos de acceso, retroalimentación en la construcción del conocimiento, utilización pedagógica de los recursos didácticos digitales, la integración de estos con las herramientas tecnológicas de los EVA, etc.

Una alternativa de solución a estos problemas, para la enseñanza de la Física, en una modalidad e-learning, donde se usan laboratorios virtuales, sería la aplicación del conocimiento pedagógico en la implementación de un diseño de aprendizaje. Este último “*implica mapear los resultados de aprendizaje deseados en actividades y experiencias de aprendizaje, teniendo en cuenta factores psicológicos, pedagógicos y tecnológicos*” (Ford, 2008).

Los factores psicológicos y pedagógicos, integrados a los tecnológicos, pueden afectar la mejor manera en que los docentes podrían presentar la información y la mejor manera de diseñar los sistemas para apoyar a los estudiantes en su aprendizaje, considerando las diferencias individuales, ya que, cada persona puede aportar al proceso de aprendizaje diferentes tipos y niveles de conocimiento previo, y también puede traer consigo motivaciones, confianza y estilos de aprendizaje muy diferentes.

En la propuesta de esta investigación el aporte se centra en el conocimiento pedagógico aplicado a los diseños de aprendizaje que realizan los docentes para el uso de los laboratorios virtuales para la enseñanza de la física; y se busca proponer un modelo de diseño de aprendizaje, que no solo soporte la enseñanza de la física, sino también permita generar un aprendizaje meta-cognitivo, autónomo y autorregulado en los alumnos universitarios.

En este trabajo se presentan los objetivos, la metodología, los resultados esperados y la formación de recursos humanos de la investigación.

2. OBJETIVOS

Con esta investigación se busca realizar contribuciones teóricas y metodológicas en el campo de la Informática Educativa con el uso de los Laboratorios Virtuales para la enseñanza de la Física, en una modalidad e-learning. En consecuencia, los objetivos son:

Objetivo General

Proponer un modelo de Diseño de Aprendizaje con el uso de Laboratorios Virtuales que no solo soporte la enseñanza de la Física en el nivel universitario, sino también permita, generar un aprendizaje meta-cognitivo, autónomo y autorregulado en los alumnos, en una modalidad e-learning.

Objetivos Específicos

1. Desarrollar el estado del Arte de cómo se usan los Laboratorios Virtuales para la enseñanza de la Física en el nivel universitario y como se evalúa el uso de los mismos.
2. Caracterizar el uso actual de los laboratorios virtuales en la enseñanza de la Física en el nivel universitario en una modalidad de e-learning.
3. Establecer pautas basadas en el diseño de aprendizaje y el conocimiento pedagógico para la integración de los Laboratorios Virtuales para la enseñanza de la Física en el nivel universitario en una modalidad e-learning, que permita un aprendizaje meta-cognitivo, autónomo y autorregulado en los alumnos.
4. Crear un modelo de Diseño de Aprendizaje para la enseñanza de la Física en el nivel universitario, con el uso de Laboratorios Virtuales, que respete las pautas establecidas.
5. Evaluar el modelo diseñado en ambientes reales.

METODOLOGÍA

Con el propósito de dar cumplimiento al objetivo específico 1, se realizarán las siguientes actividades:

- a. Exploración e investigación bibliográfica respecto a cómo se usan los Laboratorios Virtuales para la enseñanza de la Física en el nivel universitario y como se evalúa el uso de los mismos. Se indagará sobre el conocimiento pedagógico empleado en la construcción del diseño de aprendizaje.
- b. Análisis de los antecedentes encontrados.
- c. Síntesis de los antecedentes encontrados.

Con el fin de obtener el objetivo específico 2, se realizará:

- a. Trabajo de campo en las asignaturas de Física pertenecientes al departamento de Física de la FCEyT de la UNSE, relevando información sobre el uso de los laboratorios virtuales, haciendo hincapié en aspectos tecnológicos y pedagógicos de su uso. Para esto, se diseñarán los instrumentos de recolección.
- b. Caracterizar el uso de los laboratorios virtuales, a partir del relevamiento realizado y en base a la historia del arte elaborado en la etapa de exploración bibliográfica. Se identificarán y definirán las dimensiones a utilizar en la caracterización del uso de los laboratorios virtuales en la enseñanza de la Física, que tentativamente serán las vinculadas a la generación del conocimiento pedagógico, que estimulen un aprendizaje metacognitivo, autónomo y autorregulado.

Para el objetivo específico 3 se plantean las siguientes actividades:

- a. Definir aspectos necesarios y relevantes para que un aprendizaje sea metacognitivo, autónomo y autorregulado.
- b. Analizar las estrategias pedagógicas y los recursos tecnológicos en entornos de e-learning para promover el conocimiento.
- c. Definir las pautas que debería contemplar un diseño de aprendizaje que utilice laboratorios virtuales para la enseñanza de la Física en una modalidad e-learning, y genere un aprendizaje metacognitivo, autónomo y autorregulado en los alumnos.

Con el propósito de dar cumplimiento al objetivo específico 4, se realizarán las siguientes actividades:

- a. Crear el modelo de diseño de aprendizaje para la enseñanza de la Física que utilice Laboratorios Virtuales en una modalidad e-learning para el nivel universitario teniendo en cuentas las pautas definidas anteriormente.

Por último, para dar cumplimiento con el objetivo específico 5 se realizará:

- a. Definir las dimensiones de análisis a considerar en la evaluación y realizar el diseño de la evaluación.
- b. Implementar el Modelo de diseño de aprendizaje en las asignaturas de Física dependientes del departamento de Física de la FCEyT de la UNSE que fueron objeto de estudio en el trabajo de campo.

3. RESULTADOS ESPERADOS

Los resultados que se esperan obtener de esta investigación son:

- Estado del arte sobre el uso y evaluación de los laboratorios virtuales para la enseñanza de la Física en el nivel universitario.
- Caracterización del uso de los laboratorios virtuales en la enseñanza de la Física en el nivel universitario, considerando la generación de conocimiento pedagógico.
- Pautas a seguir en el diseño de aprendizaje, que utilice laboratorios virtuales para la enseñanza de la Física en una modalidad e-learning en el nivel universitario, y genere un aprendizaje metacognitivo, autónomo y autorregulado en los alumnos.
- Modelo de diseño de aprendizaje, aplicando las pautas definidas, para la enseñanza de la Física en el nivel universitario, usando laboratorios virtuales.
- Evaluación del modelo en ambientes reales de aprendizaje.

4. FORMACIÓN DE RECURSOS HUMANOS

Esta investigación permitirá la formación de recursos humanos de la Facultad de Ciencias Exactas y Tecnologías de la UNSE. En este sentido, uno de los autores desarrolla su Tesis de la carrera de Maestría en Informática Educativa de la UNSE en esta temática. Además, la investigación permitirá afianzar el grupo de investigación del IISI en temáticas tales como Diseño de aprendizaje, educación a distancia, laboratorios virtuales, etc. en la enseñanza de la Física, lo que contribuye a una mejora en el fondo de conocimiento disciplinar disponible, no sólo a nivel local sino también regional y nacional.

5. BIBLIOGRAFÍA

- Laboratorios Remotos en la Enseñanza de la Física en Latinoamérica. <https://www.researchgate.net/publication/317209671>
- Cabrera Medina, J. y Sánchez Medina, I. (2016). Laboratorios virtuales de física mediante el uso de herramientas disponibles en la Web. *Memorias De Congresos UTP*, 1(1), 49-55. Recuperado a partir de <https://revistas.utp.ac.pa/index.php/memoutp/article/view/1296>
- Da Silva, N. C. (2012). Laboratório virtual de Física Moderna: atenuação da radiação pela matéria, *Cad. Bras. Ens. Fís.*, v. 29, n. 3: p. 1206-1231, dic. 2012. Recuperado <https://periodicos.ufsc.br/index.php/fisica/article/view/2175-7941.2012v29n3p1206>
- Ford, N. *Web-based Learning Through Educational Informatics: Information Science Meets Educational Computing*. IGI. Publishing. 2008.
- López, S., Veit, E. A. y Solano Araujo, I (2016). Una revisión de literatura sobre el uso de modelación y simulación computacional para la enseñanza de la física en la educación básica y media. *Revista Brasileira de Ensino de Física*, vol. 38, nº 2, e240. <http://dx.doi.org/10.1590/1806-9126-RBEF-2015-0031>
- Ré, M.A. y Giubergia, M. F. (2013). Estudio del resorte en el laboratorio virtual. http://sedici.unlp.edu.ar/bitstream/handle/10915/27554/Documento_completo.pdf?sequence=1&isAllowed=y
- Ardura, D. y Zamora Á. (2013). ¿Son útiles entornos virtuales de aprendizaje en la enseñanza de las ciencias secundaria? Evaluación de una experiencia en la enseñanza y el aprendizaje de la Relatividad”, *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias* 11(1), 83–93. <https://revistas.uca.es/index.php/eureka/article/view/2864>
- Arguedas-Matarrita, C., Concari, S. B. y Marchisio, S. T. (2017). Una revisión sobre desarrollo y uso de Laboratorios Virtuales y

REALIDAD AUMENTADA EN ESCENARIOS EDUCATIVOS CIENTÍFICOS

Mario Bustillo, Liliana Ferrer, Silvina Videla, Gabriela Ohanian y Sergio Vardaro

Facultad de Ingeniería, Universidad Nacional de Cuyo. Mendoza, Argentina.

mabustillo02@gmail.com

lferrer@fing.uncu.edu.ar mvidela@fing.uncu.edu.ar

gohanian@fing.uncu.edu.ar svardaro@uncu.edu.ar

RESUMEN

Con el rápido crecimiento de las Tecnologías de la Información y Comunicación (TIC), se está produciendo un importante impacto en la educación. Dentro de estas, la Realidad Aumentada se destaca por adaptarse de manera eficaz a los nuevos estilos de aprendizaje demandados por los alumnos en la sociedad de la información y el conocimiento, siendo este, el punto de partida para el desarrollo de nuevas y novedosas prácticas docentes innovadoras en la educación científica. En el marco de innovación y aplicación de tecnologías disruptivas para la educación científica, en especial para el área de Química Orgánica, se desarrolla AUMENTED, plataforma educativa que permite la visualización de estructuras tridimensionales de moléculas químicas mediante el uso de realidad aumentada. A través de la implementación de estas nuevas tecnologías en la educación, estas pueden ser vistas como herramientas y materiales de construcción que promuevan, faciliten y enriquezcan la comprensión de temas y conceptos propios e importantes en el proceso de aprendizaje de las nuevas generaciones para estar a la vanguardia de los procesos educativos.

Palabras clave: Realidad Aumentada, Tecnologías Disruptivas, Educación Científica, Tecnologías de Información y Comunicación (TIC), Química Orgánica, Innovación Docente, Tecnologías Emergentes.

CONTEXTO

Esta línea de investigación y desarrollo (I/D) surge a partir de la necesidad de implementar

estrategias y herramientas TIC en la cátedra de Química Orgánica de la Facultad de Ingeniería de la Universidad Nacional de Cuyo, en el marco del proyecto “Implementación del uso de las TIC en el proceso enseñanza aprendizaje de la Química Orgánica” (segunda parte). Esto con la finalidad de mejorar el trabajo individual, la autonomía del alumnado, y la posibilidad de modificar y adaptar los métodos de evaluación para mejorar la calidad de la enseñanza.

Está claro que el objetivo que se persigue con la incorporación de estas nuevas tecnologías es mantener la atención del alumno y lograr un aprendizaje de acuerdo con el contexto tecnológico-social en el que vive, ya que pertenece a una generación que está en contacto permanente con dispositivos digitales.

Este trabajo pertenece a uno de los proyectos de investigación de tipo 1 con resolución N°4142, financiados por la Secretaría de Investigación, Internacionales y Posgrados (SIIP) de la Universidad Nacional de Cuyo (UNCuyo).

INTRODUCCIÓN

En el ámbito educativo particularmente, las Tecnologías Emergentes han generado un nuevo escenario para el desarrollo del proceso de enseñanza-aprendizaje, facilitando posibilidades múltiples y planteando nuevos desafíos en todos los niveles de formación. Actualmente, el espacio de educación formal trasciende en el aula, gracias a la existencia de múltiples herramientas disponibles tanto en dispositivos móviles como en computadoras

ocasionando la necesidad de repensar la educación. Asimismo, las Tecnologías Emergentes también permiten generar indicadores dinámicos que proporcionan nuevas métricas para el análisis del proceso de aprendizaje y la gestión educativa. (Russo et al., 2021).

Como menciona Ferrer et al. (2019-2021), las Tecnologías de Información y Comunicación (TIC) han incidido mucho en la mejora de la productividad en general. Este cambio ha obligado a las universidades a crear nuevas modalidades de estudio y a establecer formas didácticas innovadoras para hacer más comprensibles los conocimientos tomando en cuenta las diversidades de la población. Asimismo hay que tener presente que las nuevas generaciones son individuos con otros intereses de motivación y patrones de formación, y que la utilización de las tecnologías didácticas como medios educativos pueden aprovecharse como elementos motivantes para el aprendizaje, considerando la facilidad de interacción de los estudiantes con la tecnología actual.

Por otro lado, el uso de la Realidad Aumentada (RA) en el proceso de enseñanza-aprendizaje de la Química Orgánica es considerado una innovación didáctica en este tipo de contenido científico-tecnológico, además de una oportunidad para las nuevas metodologías de docencia mediante la incorporación de las TIC. La RA es un medio interactivo que añade objetos virtuales a la realidad del usuario, a través de una cámara y observado por medio de la pantalla de la computadora o dispositivo móvil, permitiendo ampliar información de un tema específico. (Ramos Geliz et al., 2015).

Punto de partida para el desarrollo de Realidad Aumentada en escenarios educativos científicos

Tal como expresa Cerillo (2020), la enseñanza de la química enfrenta una serie de retos como la evaluación educativa y la didáctica. Dentro de esta última, se localizan las acciones prácticas, las cuales van de los contenidos a la praxis del estudiante. Es aquí donde la RA podría mejorar el acercamiento de los jóvenes

a la construcción de un pensamiento científico y, sobre todo, la formación universitaria hacia las ciencias duras en un contexto de innovación tecnológica educativa, sin olvidar el enfoque científico del proyecto hacia las carreras de Ingeniería.

Además, durante el proceso de enseñanza de la Química Orgánica, se comprobó que los estudiantes presentaban dificultad con ciertos temas básicos que son fundamentales para poder llevar la continuidad de la materia. Uno de los problemas detectados fue la complicada visualización o imaginación de las estructuras químicas de las moléculas, lo cual es fundamental para determinar propiedades físicas-químicas e información relevante sobre la molécula.

Mediante el proyecto, se pretende lograr la incorporación de la aplicación de la tecnología de RA a la química para esclarecer la interacción de los elementos y sus aplicaciones, con el fin de generar talleres que complementen el aprendizaje. De la misma manera, generar un aprendizaje tecnológico que ayude a los alumnos a familiarizarse con la tecnología e informática en general, así como generar experiencias de aprendizajes a los estudiantes. Se trata de llevar el conocimiento desde una manera más clara, entretenida y fácil de visualizar; para generar una mayor comprensión de los contenidos.

Al igual que Ramos Geliz et al. (2015), el proyecto incorpora las tecnologías de RA en el campo educativo, ofreciendo un valor agregado que permite generar y captar el interés por parte de los estudiantes hacia las áreas del saber, de esta forma las asignaturas en las cuales un estudiante tenga dificultades de aprendizaje, se tornarán interactivas, facilitando así la comprensión y aprendizaje de un tema particular.

A partir de estos aspectos, el enfoque del trabajo es destinado al proceso de diseño e implementación de tecnologías disruptivas enfocadas a la enseñanza de la Química enriquecidas con el uso de la RA, destinado al desarrollo del proceso de aprendizaje en la Química dirigido a las carreras de Ingeniería.

Implementación de Realidad Aumentada

Como lo menciona Martínez-Hung (2017), en la enseñanza de la Química universitaria son comúnmente empleados los programas de modelación molecular y las bases de datos. Por otra parte, se encuentran relativamente pocas aplicaciones de la RA en la enseñanza de la Química en comparación con otras áreas del conocimiento. Quizás esto último se deba a que resulta poco visible, en la bibliografía revisada, un vínculo entre lo ampliamente conocido (programas de modelación molecular y las bases de datos) y la tecnología de RA a pesar de las ventajas demostradas en otras áreas.

Luego de una serie de pruebas con distintas plataformas existentes en el momento, las cuales permitían el uso de RA, se decidió crear una plataforma que incluyera los modelos tridimensionales de las moléculas químicas orgánicas que puedan representarse empleando la RA. El tipo de plataforma desarrollada es de estilo WebAR, que consiste en un navegador web que da acceso a la RA. De esta manera, para la interacción con la plataforma no se requiere instalar ninguna aplicación o programa adicional para su uso. Además, posee una interfaz de sencilla comprensión e intuitiva para el usuario.

La integración de este tipo de plataforma fue un desafío y una oportunidad que permitió presentar al estudiante contenidos altamente interactivos que responden a sus expectativas y necesidades con el fin de que puedan interpretar los contenidos, relacionarlos con el mundo real y evolucionar de la visualización y uso de información desde contextos en 2D (p. ej. Libros de texto) hacia uno de 3D (p. ej. Manipulación, interacción, perspectiva, complejidad, integración, etc.), construyendo así puentes entre la teoría y la experiencia práctica (modelización). (Merino, 2014).

Entorno a la aplicación de la tecnología de RA como una nueva tecnología que puede incorporarse en el proceso de enseñanza y aprendizaje dentro del sistema educativo, dentro del área de Química Orgánica, se plantea AUMENTED-Realidad Aumentada,

una plataforma educativa desarrollada por el equipo de investigación para el dictado de la clase de Química Orgánica, la cual permite la visualización de estructuras tridimensionales de moléculas químicas mediante el uso de RA. Esto con la finalidad de facilitar el aprendizaje de conceptos químicos y generar experiencias de aprendizaje significativas en los estudiantes.



A) Molécula de Dimetilacetamida



B) Molécula del Ibuprofeno

Figura 1. Visualización de moléculas con la plataforma de AUMENTED

Procedimiento para el desarrollo e implementación de experiencia incorporando Realidad Aumentada

La ejecución e implementación del trabajo de investigación se dividió en siete fases:

1. Selección de las clases donde se aplicará la RA.
2. Selección de los temas y tipologías que se incluirán con la RA. Selección de los objetos a representar con RA.
3. Construcción de los objetos tridimensionales y conversión a la plataforma de RA.
4. Diseño de actividades evaluativas y rúbricas con RA.
5. Instrucción y tutoría del uso de la plataforma de RA.

6. Aplicación de la RA en el proceso docente.
7. Validación de la aplicación mediante encuestas a los estudiantes.

Sin embargo la inclusión de las TIC en la educación debe ir acompañada de una serie de lineamientos que definan un marco de referencia para la toma de decisiones respecto de las acciones que se deben realizar durante el proceso (Hernandez, 2017). Para ello deben ir acompañadas de una adecuada fundamentación que brinde marco pedagógico, por lo que fue necesario plantear rúbricas de evaluación para las actividades que involucraron el uso de RA.

LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Los principales ejes temáticos que se están investigando son los siguientes:

1. Realidad Aumentada aplicada a la Química Orgánica

Se encuentra en investigación y desarrollo la realidad aumentada aplicada a la Química Orgánica, en la búsqueda de mostrar, a través de ella, la estructura tridimensional de las moléculas. Este instrumento funciona combinando ambientes reales con información digital, lo cual se visualiza en una pantalla con la intención de ampliar los sentidos, permitiendo así innovar en las metodologías de enseñanza y aprendizaje de la Química.

2. Implementación de TIC en el proceso de enseñanza aprendizaje de Química Orgánica

Se indaga sobre la implementación de estrategias y herramientas de Tecnologías de Información y Comunicación que contribuyan a innovar y a elevar la calidad del proceso educativo de la asignatura Química Orgánica en relación a la enseñanza tradicional en esa asignatura de las carreras de Ingeniería. La aplicación de estrategias didácticas basadas en las TIC contribuye significativamente al aprendizaje de los contenidos de Química Orgánica.

OBJETIVOS TEÓRICOS Y EXPERIMENTALES

Los objetivos en curso de las líneas de investigación presentadas son:

- Desarrollar una plataforma basada en tecnologías de Realidad Aumentada para contribuir de manera significativa en el aprendizaje de las estructuras químicas de algunas moléculas de interés en el dictado de la clase de Química Orgánica.
- Ofrecer una herramienta de innovación en los métodos de enseñanza, en los materiales educativos y en la evaluación, impulsados por las Tecnologías de la Información y Comunicación.
- Desarrollar una herramienta auxiliar en la metodología de enseñanza de la materia capaz de aplicar el conocimiento adquirido durante el transcurso del cursado (Actividades dentro de los Trabajos Prácticos).

RESULTADOS OBTENIDOS

La RA es una tecnología con un futuro prometedor, dado que posee la capacidad de enriquecer elementos de la realidad con información detallada pudiendo aplicarse en ámbitos muy variados. En el ámbito educativo puede utilizarse para complementar los materiales didácticos con modelos virtuales que estimulen la percepción y ayuden a la comprensión de los conceptos.

Para poder evaluar y validar el proyecto, al finalizar el cursado se aplicó una rúbrica de evaluación de las tecnologías implementadas durante el cursado en contexto de virtualidad, una de ellas la de RA. La actividad fue de carácter individual. Dado los resultados en las encuestas de validación de las tecnologías que se implementaron con los estudiantes que cursaron la clase de Química Orgánica en el segundo semestre del año lectivo 2020 y 2021, se observa cómo los estudiantes tomaron una postura atractiva sobre el objeto de RA y se evidencia la incidencia de captar el interés del estudiante.

Sin embargo, dados los resultados de las encuestas, se observó que durante el 2020 el 25,2% de los estudiantes consideró que respecto a las actividades pedagógicas en las que se implementó la tecnología de RA fue del todo conveniente para fijar conocimientos de manera eficiente. De manera que para el próximo semestre en que se dictó la materia en el 2021, se replanteó una nueva actividad evaluativa que incluyera la tecnología de RA y que de la misma manera cumpliera los lineamientos de evaluación de la asignatura. Se observó una mejora en el porcentaje de aceptación de los estudiantes, obteniéndose un 36,4% de aceptación por parte de los estudiantes.

Al ser la implementación de RA reciente en el campo de las ciencias, es un perfeccionamiento y mejora continua en base a la información extraída del Data Analytics que proporcionan los estudiantes para mejorar la experiencia de usuario e interfaz de usuario (UX/UI) de la implementación de esta tecnología en el área de Química.

FORMACIÓN DE RECURSOS HUMANOS

En lo concerniente a la formación de recursos humanos del proyecto, se detalla la formación de 1 estudiante de grado de Ingeniería en Mecatrónica en las líneas de I/D en Realidad Aumentada y tecnologías disruptivas aplicadas a la Química.

BIBLIOGRAFÍA

- Cerillo, S. R. (2020). Realidad aumentada y aprendizaje en la química orgánica. *Revista Apertura*, 12(1), [pp.] 106-117. [Online]. <http://dx.doi.org/10.32870/Ap.v12n1.1853>
- Ferrer, L., Videla, M. S., Ohanian., G., Hoffmann, M. L., Vardaro Páscolo, S. A., Luschini, R. F., Bustillo López, M. F. (2019-2021). "Uso de las TIC como apoyo al proceso enseñanza aprendizaje de Química Orgánica: (segunda parte)". "Use of ICT to support the teaching process of Organic Chemistry: (second part)". Proyecto de investigación (SIIP UNCuyoB068). Mendoza, Universidad Nacional de Cuyo. Facultad de Ingeniería. [Online]. <https://bdigital.uncu.edu.ar/14605>
- Hernandez, R.M. (2017). Impacto de las TIC en la educación: Retos y Perspectivas. *Propósitos y Representaciones*, 5(1), 325 – 347. [Online]. <http://dx.doi.org/10.20511/pyr2017.v5n1.149>
- Martinez-Hung H. (2017). Modelos de Realidad Aumentada aplicados a la enseñanza de la Química en el nivel universitario. *Revista Cubana de Química*. 29(1), [pp.] 13-25. [Online] http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2224-54212017000100002&lng=es&nrm=iso
- Merino, C. (2014). Realidad aumentada para el diseño de secuencias de enseñanza-aprendizaje en química. *Educación Química* 26(2), [pp.] 94-99. [Online]. <http://www.revistas.unam.mx/index.php/req/article/view/52916>
- Ramos Geliz, F., Toscano Ricardo, A., Regino Vidal, C., & Galván Lozano, E. E. (2015). Objeto virtual de aprendizaje para la enseñanza de la química del carbono soportado en dispositivos móviles y realidad aumentada. *Repositorio Digital de Universidad Nacional Autónoma de México*. [Online]. <https://repositorial.cuaed.unam.mx:8443/xmlui/handle/20.500.12579/3950>
- Russo, C. C. et al. (2021). Tecnologías emergentes. XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021, Chilecito, La Rioja). *Repositorio Institucional de la UNLP*. [Online]. <http://sedici.unlp.edu.ar/handle/10915/120955>

Desarrollo de aplicaciones para colaboración en e-learning

Rosanna Costaguta, María de los Ángeles Menini, Daniela Missio, Germán Lescano, Pablo Santana-Mansilla y Nevelin Salazar

Instituto de Investigación en Informática y Sistemas de Información (IISI)
Facultad de Ciencias Exactas y Tecnologías (FCEyT)
Universidad Nacional de Santiago del Estero (UNSE)
{rosanna, marameni, dmissio, gelescano, psantana, nsalazar}@unse.edu.ar

RESUMEN

El e-learning se refiere al uso integrado de computadoras y redes de computadoras para soportar los procesos de enseñanza y de aprendizaje. El Aprendizaje Colaborativo Soportado por Computadora (ACSC) es una forma de e-learning que surge como una evolución del software que forzaba a los estudiantes a aprender como individuos aislados. Los sistemas de ACSC proveen herramientas de software que facilitan el compartir información y conocimiento, así como también facilitan la comunicación tanto a nivel grupal como entre estudiantes y docentes. Durante las actividades colaborativas desarrolladas en estos sistemas, los estudiantes interactúan organizados en grupos para concretar las tareas asignadas, mientras son orientados y asistidos por los docentes.

Surgidos en la década de 1990, los entornos para ACSC fueron rápidamente adoptados en todos los niveles educativos debido a las ventajas que ofrecen: independizar a estudiantes y docentes de las variables tiempo y espacio, desarrollo de habilidades de liderazgo (solución de problemas, pensamiento crítico, análisis, etc.), expansión de la conciencia global en grupos con estudiantes de diferentes latitudes, entre otras. Sin embargo, a pesar de todos los beneficios potenciales del ACSC, no existen garantías de que las experiencias en ACSC sean exitosas. Hay numerosas variables que inciden en ese resultado y es aquí donde desarrollos informáticos innovadores pueden hacer la diferencia.

Esta línea de investigación tiene como propósito desarrollar aplicaciones de software que propicien prácticas colaborativas adecuadas en entornos de ACSC. Para ello, se propone que tales aplicaciones realicen el análisis de las interacciones de estudiantes y de e-tutores, utilizando técnicas de Inteligencia Artificial y de minería de datos, a fin de: a) detectar dificultades en la dinámica de colaboración de los grupos; b) reconocer las características de comportamiento de sus usuarios (estilos de aprendizaje, estilos de personalidad, o emociones manifestadas) para adaptarse a ellas; c) brindar a los estudiantes materiales generados con realidad aumentada y/u objetos de aprendizaje adaptados a sus preferencias y características individuales y/o grupales. Se intentará así promover comportamientos que beneficien tanto a los procesos de enseñanza como a los de aprendizaje, propiciando alcanzar el éxito en las experiencias de colaboración. Todos los desarrollos se validarán mediante sesiones experimentales especialmente diseñadas en las que participarán grupos de estudiantes y docentes universitarios reales. Los datos recabados producto de dichas experiencias serán procesados utilizando tanto técnicas estadísticas como métricas específicas de las técnicas utilizadas, lo cual posibilitará comprobar la validez de esta propuesta.

Palabras clave: *Aprendizaje colaborativo soportado por computadoras, estudiantes colaborativos, e-tutores colaborativos, técnicas de Inteligencia Artificial y minería de datos, emociones, personalización, realidad aumentada, objetos de aprendizaje.*

CONTEXTO

En este artículo se presenta al proyecto de investigación “*Desarrollo de aplicaciones para colaboración en e-learning*”, acreditado y financiado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Santiago del Estero (SECyT - UNSE), para ejecución en el período 2022-2025. El proyecto responde a una de las líneas prioritarias de investigación del IISI, vinculada con la creación y uso de tecnologías informáticas destinadas a mejorar los procesos de enseñanza y de aprendizaje. Por otro lado, este proyecto continúa una línea de investigación iniciada en 2010-2011 por el proyecto 23/C089 “*Fundamentos Conceptuales y Soportes Tecnológicos de la Informática Educativa*”, continuada en 2012-2016 por el proyecto 23/C097 “*Sistemas de Información Web Basados en Agentes para Promover el Aprendizaje Colaborativo Soportado por Computadora*”, y en 2017-2021 por el proyecto 23/C138 “*Mejorando escenarios de aprendizaje colaborativo soportado por computadora*”. Todos proyectos acreditados y financiados por SECyT – UNSE.

1. INTRODUCCIÓN

El término Aprendizaje Colaborativo Soportado por Computadora (ACSC) se usa para designar situaciones en las cuales el aprendizaje y la enseñanza se organizan a través de actividades colaborativas, mediadas por computadora, que involucran a grupos de estudiantes orientados y asistidos por el docente o e-tutor. La principal característica de esta modalidad es que los participantes pueden trabajar colaborativamente ubicados en puntos geográficos distantes, e incluso, contribuyendo en momentos diferentes en el tiempo.

En ACSC el concepto de grupo es fundamental. Un grupo colaborativo se define como un conjunto dinámico de estudiantes que trabajan juntos para alcanzar eventualmente alguna meta prefijada [Del2001]. Cada estudiante es responsable por sus comportamientos, pero todos trabajan juntos sobre el mismo problema respetando las

habilidades y contribuciones de cada uno. A través de sus comportamientos los estudiantes manifiestan ciertas variables que influyen en el éxito de las experiencias de ACSC: estilos de aprendizaje, estilos de personalidad, conocimientos, estados emocionales, etc. Analizar los comportamientos y adaptar los sistemas a estas podría propiciar los aprendizajes y mejorar los desempeños individuales y grupales.

Varias investigaciones realizadas en ciencias de la computación, neurociencia, educación y psicología demostraron que la cognición y la emoción, están intrínsecamente relacionadas [Fra2010, Cia2003, Ekm2004, Kor2001], por ende, pueden influir sobre el desempeño de una persona [Dam1994] y juegan un rol esencial en la atención y en la memoria, en el juicio, en la toma de decisiones y en la resolución de problemas creativos [Ise2000]. Las emociones tienen un rol importante en la interacción entre las personas, por lo tanto, es esencial reconocer cómo éstas se manifiestan en los estudiantes colaborativos y analizar su impacto sobre el aprendizaje del grupo.

Por otra parte, en el contexto del ACSC se requiere que el docente o e-tutor actúe como un moderador o facilitador que promueve y coordina las discusiones grupales, de modo tal de alcanzar los resultados de aprendizaje esperados [San2016]. Dado que las interacciones que promueven el aprendizaje no ocurren espontáneamente, es imprescindible la participación de los e-tutores para fomentar las interacciones y resolver los conflictos grupales que pudieran producirse.

El concepto de Realidad Aumentada (RA) surge en el año 1960, y hace referencia a un sistema interactivo que tiene como entrada la información del mundo real y superpone a esa realidad, nueva información digital en tiempo real. Por lo general, la RA se implementa mediante diferentes dispositivos móviles, lo que facilita la incorporación en e-learning debido al amplio uso de celulares por parte de los estudiantes [Alm2019]. Esto hace que la inclusión de RA sea una alternativa promisoriosa en ámbitos educativos. Estos contextos

abarcan al ACSC, aunque actualmente las experiencias documentadas de uso de RA en ACSC sean escasas.

Los Recursos Educativos Abiertos (REA), en adelante objetos de aprendizaje, fueron definidos por la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [Une2002] como materiales en formato digital que se ofrecen de manera gratuita y abierta a docentes, investigadores y estudiantes, es decir, material digital de dominio público o con una licencia abierta, que puede ser tanto utilizado como modificado por terceros. Los antecedentes de uso de estos tipos de recursos son numerosos, pero al mismo tiempo, son escasos en ACSC.

En este proyecto se propone investigar teórica, metodológica y experimentalmente las contribuciones que puedan realizarse en el desarrollo de aplicaciones para ACSC que incluyan técnicas de Inteligencia Artificial, técnicas de minería de datos, realidad aumentada y/u objetos de aprendizaje para promover un ACSC exitoso. El desarrollo de estas contribuciones permitirá ofrecer aplicaciones con la capacidad de adaptarse a las preferencias o inclinaciones de los estudiantes y a las habilidades de los docentes, e incluso capaces de monitorear las interacciones por ellos generadas.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO/OBJETIVOS

Para el presente proyecto de investigación se establecieron los siguientes objetivos generales:

- Generar conocimiento científico-tecnológico de relevancia, para el desarrollo de sistemas en el área del ACSC.
- Producir propuestas metodológicas y desarrollos en ACSC, que usen técnicas de Inteligencia Artificial, de minería de datos, realidad aumentada y/u objetos de aprendizaje, a fin de promover los procesos de enseñanza y de aprendizaje.

- Transferir conocimientos y ofrecer servicios al medio a través del asesoramiento y la capacitación de estudiantes de grado y postgrado, y también por medio del dictado de cursos y seminarios.
- Difundir los resultados obtenidos mediante publicaciones en revistas especializadas y de presentaciones en congresos, simposios y jornadas nacionales e internacionales.
- Formar recursos humanos mediante la realización de tesis de grado y de postgrado, y la dirección de becarios e investigadores en formación.

La pregunta central que guiará el desarrollo de la investigación es la siguiente: *¿Es posible mejorar el desempeño académico de los estudiantes universitarios y el desempeño de los docentes, en los sistemas de ACSC, mediante el uso de técnicas de Inteligencia Artificial y minería de datos, realidad aumentada y objetos de aprendizaje?*

Acorde con este interrogante, se fijaron los objetivos específicos que se enuncian a continuación:

- Determinar el estado actual de conocimiento y desarrollo de sistemas que utilicen técnicas de Inteligencia Artificial y de minería de datos, realidad aumentada y objetos de aprendizaje en el ámbito del ACSC.
- Desarrollar aplicaciones que utilicen técnicas de Inteligencia Artificial y de minería de datos, realidad aumentada y/u objetos de aprendizaje, considerando, por ejemplo, características de los estudiantes tales como: habilidades de colaboración, roles de grupo, estilos de aprendizaje, estilos de personalidad y estilos emocionales.
- Desarrollar aplicaciones que utilicen técnicas de Inteligencia Artificial y de minería de datos, realidad aumentada y/u objetos de aprendizaje, considerando, por ejemplo, las habilidades de los

docentes que se desempeñan como e-tutores.

- Evaluar el impacto producido por la inclusión de técnicas de Inteligencia Artificial y de minería de datos, realidad aumentada y/u objetos de aprendizaje, en los procesos de enseñanza y de aprendizaje, en el ámbito de los sistemas de ACSC en contextos universitarios.

A partir del año 2010, a través de los distintos proyectos de investigación mencionados en la sección Contexto, este equipo identifica y analiza factores y situaciones que influyen en el ACSC. Algunos de los resultados del grupo de investigación que servirán de base para alcanzar los objetivos de este nuevo proyecto se enuncian en [Cos2008, Cos2009, Cos2011, Cos2013, Cos2014, Cos2015, San2016, Les2016, Les2020a, Les2020b, Les2021].

3. RESULTADOS ESPERADOS

Con la concreción de este proyecto se esperan obtener los siguientes resultados principales:

- Estado del arte vinculado con el uso de técnicas de Inteligencia Artificial y de minería de datos, realidad aumentada y/u objetos de aprendizaje, en sistemas de ACSC
- Identificación de características que impactan en el desempeño los estudiantes a ser consideradas en sistemas de ACSC
- Identificación de características que impactan en el desempeño de los docentes a ser consideradas en sistemas de ACSC.
- Diseño de aplicaciones que incorporen técnicas de Inteligencia Artificial y de minería de datos, realidad aumentada y/u objetos de aprendizaje, para mejorar el desempeño de los estudiantes en sistemas de ACSC.
- Diseño de aplicaciones que incorporen técnicas de Inteligencia Artificial y de minería de datos, realidad aumentada y/u

objetos de aprendizaje, para mejorar el desempeño de los docentes en sistemas de ACSC.

- Despliegue de las aplicaciones desarrolladas.
- Ejecución de actividades experimentales que permitan validar el correcto funcionamiento de las aplicaciones desarrolladas.
- Presentación de resultados en eventos científicos nacionales e internacionales y publicaciones en revistas indexadas.

Por otra parte, el desarrollo del proyecto continuará consolidando el trabajo de este equipo de investigación en el área del ACSC.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo de este proyecto está constituido por un docente investigador formado, cinco en formación, dos estudiantes de posgrado y cinco estudiantes de grado. En el marco del proyecto se desarrollarán: 1 tesis doctoral, 2 de maestría, 3 de especialización y 5 de licenciatura.

5. BIBLIOGRAFÍA

- [Alm2019] Almenara J., Osuna J. & Llorente, C. (2019). *La realidad aumentada en la enseñanza universitaria*. Revista de Docencia Universitaria REDU, Vol.17, pp. 105-118
- [Cia2003] Cialdini R. (2003). *Influence: Science and Practice*. Allyn Bacon.
- [Cos2008] Costaguta R. & Amandi A. 2008. *Training collaboration skills to improve group dynamics*, Proceedings ACM of the 2008 Euro American Conference on Telematics and Information Systems, Aracajú, Brasil.
- [Cos2009] Costaguta R., Durán E., y Gola M. 2009. "El modelo B-learning y el aprendizaje colaborativo en una experiencia innovadora de enseñanza-aprendizaje". VI Encuentro Nacional "La Universidad como Objeto de Investigación". Córdoba, Argentina.

- [Cos2011] Costaguta R., García P. y Amandi A. 2011. *Entrenando las habilidades de colaboración de los estudiantes mediante agentes*. IEEE Latin America Transactions. Vol. 9(7), pp. 1118-1124.
- [Cos2013] Costaguta R., Menini M., Missio D., Santana Mansilla P. y Yanacón Atía D. 2013. *Modelos Multiagentes Recomendadores Adaptativos en Grupos*. V Congreso Internacional de Ambientes Virtuales de Aprendizaje Adaptativo y Accesible, 2013.
- [Cos2014] Costaguta R. & Menini M. 2014. *An Assistant Agent for Group Formation in CSCL based on Student Learning Styles*. 7th Euro American Conference on Telematics and Information Systems (EATIS 2014), Valparaiso, Chile.
- [Cos2015] Costaguta R., Menini M., Missio D., Santana Mansilla P., Lescano G. y Yanacón Atía D. 2015. *Potenciando el Aprendizaje Colaborativo Soportado por Computadora: Algunas ideas traducidas en acciones*. XVI Virtual Educa, Guadalajara, México.
- [Dam1994] Damasio A. R. (1994). *Descartes' error: emotion, reason, and the human brain*. Grosset/Putnam Publications.
- [Del2001] Delgado A., Olguín C. y Ricarte I. (2001). *Monitoring Learners Activities in a Collaborative Environment*. Proceedings of the 7th IEEE International Workshop on Groupware. Germany.
- [Ekm2004] Ekman P. (2004). *Emotions Revealed: Recognizing Faces & Feelings to Improve Communication and Emotional Life*. Henry Holt & Co Publications.
- [Fras2010] Frasson C. & Chalfoun P. (2010). *Managing Learner's Affective States in Intelligent Tutoring Systems*. Advances in Intelligent Tutoring Systems, Vol. 308, pp. 339-358.
- [Ise2000] Isen A. M. (2000). *Positive Affect and Decision Making*(ed). Handbook of Emotions, 2da Edición, Guilford Publications, pp. 417-435.
- [Kor2001] Kort B., Reilly R. & Picard R. (2001). *An affective model of interplay between emotions and learning: Reengineering educational pedagogy—building a learning companion*. Proceedings IEEE International Conference on Advanced Learning Technology. pp. 43-48.
- [Les2016] Lescano G., Costaguta R. & Menini M. (2016). *Applying Data Mining to Discover Successful Collaborative Groups Styles*. 8th Euro American Conference on Telematics and Information Systems (EATIS 2016), Cartagena de Indias, Colombia.
- [Les2020a] Lescano G., Lara C., Collazos C. & Costaguta R. (2020). *A Technique for Conflict Detection in Collaborative Learning Environment by Using Text Sentiment*. LNAI 12469(II), pp. 39-50.
- [Les2020b] Lescano G., Costaguta R. & Amandi A. (2020). *Reconocimiento de emociones en texto*. Capítulo 4 en: *Introducción a la Computación Afectiva*, pp 55-95. Editores: José Aguilar, John Edgar Amaya, Ángel Gil. Editorial FEUNE.
- [Les2021] Lescano G., Torres-Jiménez J., Costaguta R., Amandi A. & Lara-Alvarez C. (2021). *Detecting conflicts in Collaborative Learning through the Valence Change of Atomic Interactions*. Expert System with Application, Vol.183(30).
- [San2016] Santana-Mansilla P., Costaguta R. & Schiafino S. (2016). *Towards E-Tutors Training in On-Line Collaborative Learning*. Proceedings of 9th Euro American Conference on Telematics and Information Systems (EATIS 2016)
- [Une2002] UNESCO (2002). *Forum on the Impact of Open Courseware for Higher Education in Developing Countries - Reporte Final*, UNESCO: Paris. Disponible en: CI.2002/CONF.803/CLD

Construcción de una herramienta de soporte para la creación de diagramas de clases destinados a estudiantes con disminución visual.

Ing. Clarisa Liliana Stefanich^{1,2}, Nicolas Gabriel del Valle^{1,3}, Ing. Laura del Carmen Ligorria,^{1,4}
Ing. Roxana María Manera^{1,5}, Ing. Lorena Daniela Peralta^{1,6}, Ing. Alejandra Di Gionantonio^{1,7},
Juan Gabriel Moreno^{1,8}

¹UTN Facultad Regional Córdoba

{²clarystefanich, ³dv.nico13, ⁴liuniversidad, ⁵roxanamanera, ⁶peralta.lorena.d, ⁷
ing.alejandradg}@gmail.com, ⁸gabriel_2008@live.com.ar

RESUMEN

En los últimos tiempos se ha observado la necesidad de accesibilidad de estudiantes con disminución visual total o parcial a la información concerniente a diagramas visuales llevados a cabo en cátedras de análisis, diseño y programación de software en carreras de ingeniería, tecnicatura y cursos orientados a la informática. Teniendo en cuenta la importancia de los diagramas UML en el cursado de estas asignaturas, se plantea la construcción de una herramienta de software que permita la generación y lectura de diagramas UML por medio de un lenguaje textual formalizado, y que a su vez brinde la posibilidad de interacción por parte del docente y el resto de los estudiantes mediante la manipulación de elementos gráficos (formularios, elementos visuales del diagrama), favoreciendo la construcción de conocimientos colaborativos entre estudiantes y docentes.

Palabras claves: UML, visibilidad reducida, diagrama de clase, accesibilidad.

CONTEXTO

El presente trabajo se realiza en el Área de Investigación del Departamento de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional - Facultad Regional Córdoba.

1. INTRODUCCIÓN

En relación a las herramientas educativas actuales con respecto a la accesibilidad y adaptabilidad en los diferentes niveles de educación, existen esfuerzos enfocados a la

educación preescolar, primaria y secundaria, pero no así en el nivel superior y terciario.

En el ámbito legal y de estándares, se detectan avances en la accesibilidad de la información y del medio físico tanto a nivel nacional como internacional [1][2].

En relación a los antecedentes relevados, se observa la presencia de soluciones propuestas por diferentes universidades y organizaciones, las cuales se describen brevemente a continuación:

En [3] los autores desarrollaron la aplicación “TeDUB”, la cual traduce el contenido de dibujos técnicos (electrónica, arquitectura e ingeniería de software) en descripciones mapeadas jerárquicamente, en formato XML, las cuales son accedidas por dispositivos de salida, voz, efectos de sonido (2D y 3D), Braille y dispositivos hápticos. El usuario puede interactuar con el sistema y puede explorar la jerarquía del dibujo técnico usando el teclado o un joystick. También se puede usar una pantalla braille para navegar.

En [4] los autores crearon el software “Audible Browser”, la cual interpreta archivos XMI 2.1, especialmente de diagramas de clases y estados. Luego el programa representa los nodos del diagrama en grupos de tabs. Un grupo de tabs es un conjunto de ítems comunes que pueden ser leídos usando las teclas de arriba y abajo. Los grupos de tabs para una clase son el nombre, atributos, operaciones y asociaciones, permitiendo diferentes niveles de navegación, desde lo más general a los más específico. Para transmitir la estructura de un diagrama, el software lo presenta utilizando sonidos no verbales en combinación con audio verbal. El audio no verbal se transmite más a la

izquierda o derecha del auricular representando la posición “X” del elemento en el diagrama y el tono representa la posición “Y”.

En [5] los autores generaron “Prisca”, una aplicación que permite transformar los elementos de un diagrama UML a un vista 3D, para luego ser visualizada por medio de una impresora 3D. Los textos son impresos utilizando Braille.

Por otro lado, la página web yUML[6], permite la generación de diversos diagramas UML por medio de un lenguaje textual.

En las herramientas observadas, se visualizan diversas áreas de mejora:

- Disponibilidad offline, permitiendo usar la herramienta sin necesidad de Internet.
- Interacción transparente entre docente y estudiante, manteniendo a cada uno en su contexto de trabajo/conocimiento. Esto implica que la herramienta provee dos vistas simultáneas del mismo diagrama, optimizadas cada una para el docente y el estudiante respectivamente.
- Lenguaje textual compatible con los sintetizadores de voz en español, disponibles en los lectores de pantalla.
- Permitir la creación y edición de diagramas desde cualquiera de las dos vistas: docentes o estudiantes.
- Multiplataforma.
- Creación y edición de diagramas UML de clases, casos de uso, estados y secuencia. Los cuales son los más usados en ámbitos de formación superior.

Estas áreas serán abordadas por el presente trabajo en los siguientes apartados.

2. OBJETIVOS

El objetivo de este trabajo de investigación es: “La construcción de una herramienta de software que permita la conversión de un lenguaje formalizado a un diagrama gráfico y su transformación inversa para poder ser interpretados por estudiantes con disminución visual total o parcial por medio de un lector de pantalla”.

3. HERRAMIENTAS

Para la creación del módulo de renderizado de la herramienta se utilizó la librería React [7], la cual permite administrar los eventos y el DOM correspondientes a un entorno web. Esto posibilita el empleo de múltiples componentes gráficos probados.

El software está dividido en módulos cohesivos, es decir que los contenidos dentro del mismo están relacionados semánticamente. Cada módulo es flexible en cuanto a la implementación de los algoritmos necesarios para lograr los objetivos del mismo. Esto se logró por medio del uso de la composición de objetos y el patrón estrategia.

Para la construcción de la sintaxis empleada por el estudiante para la generación de diagramas UML, se utilizó como base una gramática libre de contexto, un analizador sintáctico y una definición de la semántica del lenguaje. Dichos conceptos fueron definidos por medio de la librería Ohm [8].

Al emplear React en la construcción de la aplicación, se dispone de la API Canvas, la cual permite la creación y gestión de elementos gráficos. Para facilitar la generación visual de diagramas UML, se empleó la librería Konva [9], la cual es una capa de abstracción sobre Canvas, ocultando la complejidad inherente de dicha API.

HTML y los navegadores web poseen mecanismos maduros de sincronización para con los lectores de pantalla. Con la finalidad de utilizar esta ventaja y para satisfacer el requerimiento de aplicación offline, se empleó la librería Electron [10], la cual incluye funcionalidades que habilitan la ejecución de aplicaciones construidas con tecnología web para ser ejecutadas en un entorno de escritorio. El software queda encapsulado dentro de un navegador minimalista (chromium) y visualizado por medio de un motor de renderizado. La librería anterior satisface el requerimiento de multiplataforma, ya que permite la creación de ejecutables de los sistemas operativos más utilizados (Windows, Linux y MAC OS).

4. ESTADO DE AVANCE

El equipo implementó una herramienta de software que posibilita la conversión bidireccional simultánea entre una definición en lenguaje estructurado y un diagrama gráfico de clases, permitiendo de esta manera la interacción entre un estudiante y un docente, o para los estudiantes entre sí, en un entorno académico.

La interfaz cuenta con dos vistas, una corresponde al estudiante con disminución visual y la otra corresponde al docente o a otro estudiante. Esto fue concebido con miras a separar los ámbitos de interacción de forma tal que el estudiante no necesite interactuar con elementos visuales, y el docente o compañero de grupo no requiere aprender el lenguaje formal de generación de diagramas.

Con respecto a la vista del estudiante con disminución visual, la misma cuenta con dos sub-vistas, una corresponde a una descripción detallada de los elementos del modelo, y la otra corresponde a la definición sólo de las relaciones entre los mismos. Esta última sub-vista permite al estudiante realizar un análisis de grano grueso del diagrama (nombre de entidades y relaciones entre ellos).

El estudiante puede crear, modificar o borrar elementos del diagrama por medio de un lenguaje formal. El mismo está en español y fue construido teniendo como foco la correcta interpretación del mismo por parte de un lector de pantalla con configuración genérica en español.

En referencia a la vista del docente o compañero de grupo, se visualiza el diagrama construido y por medio de formularios e interacción con los elementos visuales del diagrama, se puede crear, editar o borrar los componentes de dicho diagrama.

Ambas vistas mencionadas están sincronizadas entre sí, teniendo como base el vector de elementos creados y sus propiedades.

Con fines a visualizar la interacción entre un docente/compañero de grupo y un estudiante con disminución visual, se expone a continuación el siguiente enunciado:

Ejemplo de aplicación

El docente plantea la realización de un diagrama de clases sobre la funcionalidad “Generación de préstamos de libros” correspondiente al dominio “Biblioteca”.

El estudiante crea el diagrama solicitado por medio del lenguaje, como muestra la figura 1:



Figura 1. Lenguaje de generación de diagramas.

El docente visualiza el modelo gráfico elaborado por el estudiante, el cual se observa en la figura 2:

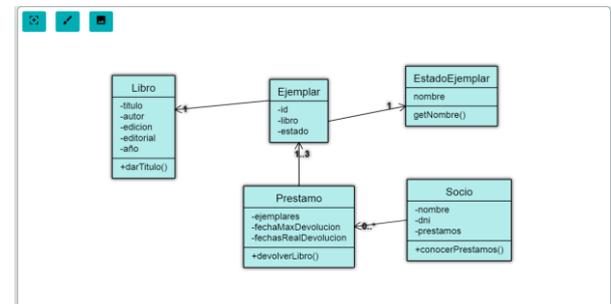


Figura 2. Diagrama de clases generado.

El docente analiza el modelo, y concluye que es necesario de acuerdo al dominio, la inclusión de una nueva clase llamada “Estado Socio”. La misma es generada por medio de la interfaz de la herramienta, sin necesidad de conocer el lenguaje de generación de diagramas utilizado por el estudiante previamente. En las figuras 3 y 4 se observa cómo el docente crea esta nueva clase y su relación correspondiente:

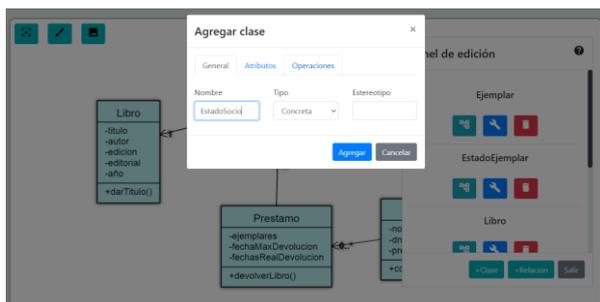


Figura 3. Creación clase “EstadoSocio”.



Figura 4. Creación relación “Socio” a “EstadoSocio”.

El diagrama gráfico se actualiza con las modificaciones realizadas por el docente, reflejadas en la figura 5:

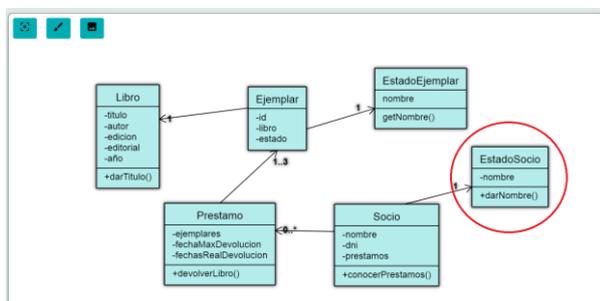


Figura 5. Diagrama con nueva clase.

A su vez, esta modificación se sincroniza con el lenguaje textual percibido por el estudiante, apreciado en la figura 6:

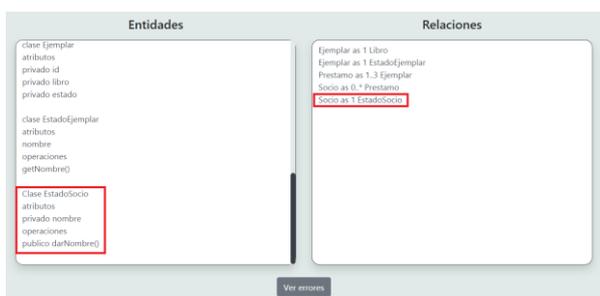


Figura 6. Lenguaje textual actualizado.

5. RESULTADOS Y AVANCES

Considerando la interacción del docente, la misma ha sido validada por medio de pruebas ejecutadas por parte de educadores de nivel superior, optimizando la usabilidad de la aplicación.

Hasta el momento el software permite tanto a docentes como a estudiantes la visualización, creación, y la modificación de diagramas de clases.

La investigación proseguirá con la realización de nuevos tipos de diagramas UML, la adaptación de la interfaz para que la misma sea compatible con las guías WCAG [11] y la validación de la herramienta por parte del estudiante.

6. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está conformado por docentes-investigadores pertenecientes a la carrera de grado de Ingeniería en Sistemas de Información.

El grupo está compuesto por una Directora, tres ingenieras investigadoras de apoyo, una ingeniera y estudiantes aspirantes a incorporarse a la carrera de investigador.

Este proyecto enriquecerá la experiencia en la carrera de investigador de los integrantes del mismo.

7. BIBLIOGRAFÍA

[1] Ley de educación superior N° 25573. Modificación ley N° 24521. (2002, 11 de Abril). [Online]. Disponible en: <http://servicios.infoleg.gob.ar/infolegInternet/anexos/70000-74999/73892/norma.htm>.

[2] Guenaga, L., Barbier, A. and Eguíluz, A. “La accesibilidad y las tecnologías en la información y la comunicación”. *Trans. Revista de Traductología*, pp. 155-169. 2007.

[3] Födisch, M., Crombie, D., & Ioannidis, G.T. “TEDUB : Providing access to technical drawings for print impaired people”. 2002.

[4] Coburn, S., Owen C. “UML Diagrams for Blind Programmers” en ASEE North Central Section Conference, 2014.

[5] Doherty B., Cheng B. “UML Modelling for Visually-Impaired Persons” en First International Workshop on Human

Factors in Modeling co-located with ACM/IEEE 18th International Conference on Model Driven Engineering Languages and Systems (MODELS 2015), Ottawa, Canada, 28 de Septiembre, 2015.

[6] yUML. [Online]. Disponible en: <https://yuml.me/>.

[7] Getting Started - ReactJS. [Online]. Disponible en: <https://reactjs.org/docs/getting-started.html>

[8] GitHub - harc/ohm: A library and language for building parsers, interpreters, compilers, etc. [Online]. Disponible en: <https://github.com/harc/ohm>

[9] Konvajs. Konvajs/Konva: Konva.js is an HTML5 Canvas javascript framework that extends the 2D context by enabling Canvas interactivity for desktop and mobile applications. [Online]. Disponible en: <https://github.com/konvajs/konva>.

[10] Electron. Electron/Electron: Build cross-platform desktop apps with JavaScript, HTML, and CSS. GitHub. [Online]. Disponible en: <https://github.com/electron/electron>

[11] W3C WAI. Web content accessibility guidelines (WCAG) overview. [Online]. Disponible en: <https://www.w3.org/WAI/standards-guidelines/wcag/>.

Recomendaciones utilizando Inteligencia Artificial a partir de las métricas de evaluación obtenidas de las analíticas de aprendizaje aplicadas a Serious Games

Leonel Domingo Guccione¹, Stella Maris Massa¹

¹Facultad de Ingeniería/ Universidad Nacional de Mar del Plata/Argentina
(7600) Av. Juan B. Justo 4302, +54-223-481660
leonel.guccione@gmail.com, smassa4@gmail.com

RESUMEN

Los Serious Games (SG) contribuyen al proceso de enseñanza/aprendizaje aportando una motivación y atractivo intrínsecos de una actividad lúdica. Permiten al estudiante/aprendiz sumergirse en un ambiente con reglas propias, que absorben toda su atención. A medida que el jugador se enfrenta con los desafíos propuestos por el juego, va dejando indicadores de su accionar, que son luego colectados mediante las analíticas de aprendizaje (Learning Analytics, LA) con la intención de evaluar el grado de avance (en el conocimiento) que el aprendiz ha alcanzado y también descubrir cuáles aspectos debe reforzar o aprender. Esta evaluación de las métricas obtenidas a partir de las LA, junto con el perfil del usuario y teniendo en cuenta el objetivo académico que se pretende lograr, permiten hacer una devolución (feedback, recomendación) al jugador / aprendiz en cuanto al desempeño alcanzado. Esto no depende necesariamente del grado de éxito o puntuación obtenidas en el juego.

Se pretende construir un marco teórico-práctico que permita generar un sistema de recomendaciones utilizando herramientas de Inteligencia Artificial (IA), a partir de las métricas de evaluación que arrojan las LA.

Palabras clave: Modelos de desarrollo, Analíticas de Aprendizaje, Inteligencia Artificial, Serious Games.

CONTEXTO

Este artículo presenta la investigación realizada en el marco de la tesis:

“Incorporación de Recomendaciones utilizando Inteligencia Artificial a partir de las métricas de evaluación obtenidas de las analíticas de aprendizaje aplicadas a Serious Games” de la Maestría en Ingeniería de Software de la Facultad de Informática de la Universidad de la Plata.

Se inserta además en el proyecto: “Modelos de Desarrollo de Serious Games. Las Analíticas de Aprendizaje e Inteligencia Artificial (2022/2023) del Grupo de Investigación en Tecnologías Interactivas (GTI) de la Facultad de Ingeniería de la Universidad Nacional de Mar del Plata.

INTRODUCCIÓN

En 1970 Clark Abt, en su libro *Serious Games* [1] introduce el término “Serious Games” para juegos de mesa, que proporciona una definición general aplicable a los juegos informáticos: “... un juego es una actividad entre dos o más personas con capacidad de tomar decisiones que buscan alcanzar unos objetivos dentro de un contexto limitado...Nos interesan los juegos serios porque tienen un propósito educativo explícito y cuidadosamente planeado, y porque no están pensados para ser jugados únicamente por diversión.”

Según Spinelli y Massa [2] algunos autores como Abt, Sawyer y Zyda, definen un SG como un videojuego que además de entretener, pretende incorporar un nuevo conocimiento o modificar alguna conducta en el jugador. Según De Gloria et al [3] los SG tiene como objetivo mejorar los procesos de aprendizaje proporcionando herramientas atractivas, motivadoras y efectivas que

también pueden crear situaciones positivas entre los estudiantes y con los profesores.

Según señala Bellotti, F [4] las simulaciones y los SG son un medio para adquirir de forma segura y rentable conocimientos (habilidades y actitudes) duraderos que son difíciles de conseguir usando la memoria.

Rocha, citado en [2], afirma que el éxito de estos juegos depende de una Especificación de Requerimientos que incluya los aspectos pedagógicos (contenido y tareas), la mecánica y trama del juego, los requerimientos del software, así como los mecanismos que nos permitan evaluar el aprendizaje, antes, durante y después del juego; y su eficacia pedagógica.

Dado el rol relevante del jugador en estos productos, los mismos deben ser diseñados siguiendo los principios del Diseño Centrado en el Usuario (DCU), este consiste en realizar las tareas de diseño teniendo presente las necesidades del usuario.

En este proceso de diseño, es interesante incorporar el concepto de “usabilidad pedagógica” propuesto por Massa [5]: “La facilidad de aprendizaje, eficiencia de uso pedagógico y la satisfacción con las que las personas son capaces de realizar sus tareas gracias al uso del producto con el que está interactuando”.

Todo juego serio tiene en cuenta los siguientes tres principios básicos de la gamificación [6][7]

-Mecánica del juego: reglas de funcionamiento.

-Dinámica del juego: acciones que surgen cuando los jugadores utilizan las mecánicas (recompensas, uso de sistemas competitivos o colaborativos, etc.)

Componentes de juego: implementaciones específicas de las mecánicas y dinámicas (avatares, de puntuaciones de juego, el área de juego, medallas, insignias y el estatus del jugador frente al resto de jugadores, etc.)

El diseño e implementación de las mecánicas de juego, aprendizaje y evaluación pueden funcionar en conjunto para crear una experiencia de juego que sea efectiva, tanto como experiencia de juego como experiencia

de aprendizaje / instrucción [8]. La mecánica de juego describe la actividad esencial del juego (nivelación, gestión de recursos y toma de turnos). La mecánica de aprendizaje describe actividades cuyo objetivo principal es el aprendizaje; son de naturaleza teórica y se implementan en el juego mediante el uso de la mecánica del juego. La mecánica de la evaluación se compone de actividades que tiene como objetivo principal la evaluación, por ejemplo un juego que desafía a los jugadores en forma progresiva, adaptando el nivel de dificultad según el rendimiento del jugador.

La evaluación se sirve de datos para demostrar el grado de cumplimiento de las metas y objetivos establecidos (para el aprendizaje) [9]. El aprendizaje es una construcción compleja que no es simple de medir incluso en un ambiente como una simulación o SG [10][11], ya que la propia naturaleza de un SG dificulta la recopilación de los datos.

Según Massa y Kühn [12] la evaluación del aprendizaje, tradicionalmente se ha hecho a través de cuestionarios completados por los estudiantes antes y después de interactuar con el juego. En el proyecto de investigación “El Proceso de Desarrollo de Serious Games. Modelos, Herramientas y Analíticas de Aprendizaje.”, Massa 2020, se introducen las LA para la obtención de métricas relacionadas con el desempeño (estudiante/aprendiz y docente/instructor), en los diversos contextos de aplicación. Las analíticas de aprendizaje constituyen una forma de obtener las métricas para conocer (evaluar) el aprendizaje logrado por el estudiante/jugador luego de la experiencia de juego (independientemente del resultado obtenido).

Las LA comienzan como desarrollos informáticos no relacionados con el aprendizaje, sino para satisfacer necesidades comerciales de comprensión de datos organizacionales internos a la empresa y el comportamiento del consumidor. Se basan en las Business Intelligence y Data Mining. [13]. Estos conceptos son aplicados a la educación, generando herramientas de análisis del aprendizaje. En la Primer Conferencia

Internacional sobre Análisis de Aprendizaje (2011) se define el concepto Learning Analytics: "La analítica de aprendizaje es la medición, recopilación, análisis e informe de datos sobre alumnos y sus contextos, con el fin de comprender y optimizar el aprendizaje y los entornos en los que se produce" .

Massa et al [14] señala que las LA aplicadas a SG permiten rastrear datos que muestran el comportamiento de los estudiantes en base a su interacción (proceso de juego), así como también permiten interpretar el proceso de aprendizaje, realizar recomendaciones y personalizar el aprendizaje (Johnson et al., 2012; Serrano-Laguna, et al. 2014; Siemens, 2010). También señala que para aplicar las tecnología emergentes apropiadas a la captura de datos (web-logs, motores de seguimiento, eye trackers, localización y detectores de movimiento, en combinación con las emergentes Analíticas de Aprendizaje) se puede adicionar un "rastreador" al SG que envía esta información a un servidor. El análisis de esto pueden producir información relevante sobre las interacciones de los estudiantes con el juego, haciendo que el conjunto de acciones, errores y aciertos del jugador sea significativo (Kühn, 2019; Loh, Sheng & Ifenthaler, 2015).

La retroalimentación (feedback) es una forma de abordar una mecánica de evaluación, tiene la "capacidad de convertir cada ítem de trabajo evaluado en un instrumento para el mayor desarrollo del aprendizaje de cada alumno" [15]. Es la información correctiva proporcionada por el docente / instructor hacia el estudiante / jugador en cuanto al desempeño (performance) realizado [16]. La retroalimentación es una consecuencia del desempeño. La retroalimentación apoya el rendimiento académico, promueve la motivación, la autorregulación y la autoeficacia; lo cual permite al estudiante disminuir la distancia entre su desempeño actual y el deseado [17]

La autoeficacia académica hace referencia a la opinión que el estudiante tenga sobre lo que puede o no hacer con los circunstancias

cambiantes del entorno académico [18]. La autorregulación del aprendizaje puede definirse como una manera de utilizar los propios recursos para planear, controlar y analizar la acción a realizar frente a actividades, tareas y elaboración de productos de aprendizaje.

Según Gaycho-Rodriguez [19], la retroalimentación mejora el aprendizaje (Jiménez & González, 2016; Sadler, 1989) en diferentes contextos (Brobst & Ward, 2002; Coddington, Feinburg, Dunn, & Pace, 2005), particularmente la retroalimentación del desempeño puede tener mayor utilidad (Zubiaur, Oña, & Delgado, 1999) que la retroalimentación de la evaluación.

Para que la retroalimentación esté completa debe incluir tres elementos: "Feed - U", "Feed - Back" y "Feed - Forward"; hacia dónde voy, cómo me está yendo y qué sigue ahora [20].

Con la información obtenida a través de las LA se pueden establecer las recomendaciones (retroalimentación) hacia el estudiante / jugador necesarias para que continúe con su aprendizaje. La retroalimentación determinará la forma en que el SG continuará sobre un nuevo nivel. Westera et al [21] determinan dos etapas en el análisis de los datos relacionados con el desempeño del usuario, la primera tiene como objetivo mejorar y personalizar la interacción con el sistema y la segunda, posterior al juego, tiene como objetivo adaptar el SG generando nuevas intervenciones para mejorar el aprendizaje.

Según Toledo et al [7] si un jugador pierde demasiadas veces (al jugar nuevos niveles) su motivación comenzará a decaer, con lo cual, el nuevo nivel de juego deberá contener un nivel de dificultad no muy elevado; pero si gana demasiadas veces, terminará por aburrirse. El sistema de recomendación tomará como base las habilidades que posee el usuario y los resultados obtenidos con el fin de subir o bajar la dificultad del nuevo nivel.

Es conveniente profundizar sobre el uso de la IA como herramienta para la generación del feedback hacia al estudiante / jugador, que permita la configuración del nuevo nivel que deba enfrentar el mismo. A partir del análisis

de las métricas arrojadas por las LA, el sistema de recomendaciones tendrá en cuenta las habilidades del jugador, los aciertos y errores, la motivación y el objetivo de aprendizaje, para configurar un nuevo nivel de juego que permita al estudiante ir disminuyendo la distancia entre su desempeño actual y el deseado.

LÍNEAS DE INVESTIGACIÓN

- Serious Games
- Recomendaciones a partir de LA
- Inteligencia Artificial

RESULTADOS Y OBJETIVOS

El objetivo del proyecto de tesis que se presenta en este artículo es diseñar e implementar un sistema adaptativo en los Serious Games incorporando recomendaciones mediante técnicas de Inteligencia Artificial a partir de las Analíticas de Aprendizaje. Se desarrollará un trabajo de campo para analizar la factibilidad.

FORMACIÓN DE RECURSOS HUMANOS

La formación de recursos humanos se considera un aspecto imprescindible e insoslayable. Integrantes del proyecto se encuentran desarrollando y dirigiendo seis tesis de postgrado, en el marco del proyecto de investigación, correspondientes al Doctorado en Humanidades y Artes- Mención Ciencias de la Educación (UNR); al Doctorado en Modelado y Simulación Computacional (UNMdP); a la Maestría en Ingeniería de Software (UNLP) y a la Especialización en Docencia Universitaria (UNMdP).

REFERENCIAS BIBLIOGRÁFICAS

[1] Clark, A. Serious Games. University Press of America, 1987

[2] Spinelli, A., Massa S. Elicitación de Requerimientos Centrada en el Usuario para el Desarrollo de Serious Games. XIX WICC 2017.

[3] De Gloria, A., Bellotti, F., Berta, R., Lavagnino, E. Serious Games for education and training. *Internacional Journal of Serious Games*. ISSN: 2384-8766. January 2014.

[4] Bellotti, F., Kaplalos, B., Lee, K., Moreno-Ger, P., Berta, R. Evaluación en y de juegos serios: descripción general. *Hindawi Publishing Corporation Advances in Human-Computer Interaction Volumen 2013*, ID de artículo 136864, 11 paginas.

[5] Massa, S. M. (2013). *Objetos de Aprendizaje: Metodología de Desarrollo y Evaluación de Calidad*, (Tesis Doctoral). Facultad de Informatica, UNLP, La Plata.

[6] Barriales A., Vásquez Paragulla, J. y Andrade-Arenas, L. Gamification as part of teaching and its influence on learning computational algorithms. *IEEE 2020 IEEE World Conference on Engineering Education (EDUNINE) - Bogota, Colombia (2020.3.15-2020.3.18)*.

doi:10.1109/EDUNINE48860.2020.9149510

[7] Toledo, P., González González, C., Muñoz, V., Aciar, S., Hernandez Rodriguez, M., Fardoun, H. Sistema de recomendación para Juegos Serios. *Actas del V Congreso Internacional de Videojuegos y Educación (CIVE '17)*. Tenerife, España.

[8] Kinzer, C. K., Hoffman, D., Turkay, S., Gunbas, N., Chantes, P., Dvorkin, T., & Chaiwinij, A. The Impact of Choice and Feedback on Learning, Motivation, and Performance in an Educational Video Game. 2012. In *Proceedings of the Games, Learning, and Society Conference (Vol. 2, pp. 175-181)*. Pittsburgh, PA: ETC.

[9] Chin, R. Dukes y W. Gamson, "Evaluación en simulación y juegos: una revisión de los últimos 40 años", *Simulación y juegos*, vol. 40, no. 4, págs. 553–568, 2009.

[10] RT Hays, "La efectividad de los juegos educativos: una revisión y discusión de la literatura", *Tech. Rep.2005-004*, Centro Naval División de Sistemas de Capacitación, 2005. de Guerra Aérea,

[11] G. Bente y J. Breuer, "Haciendo explícito lo implícito: medición incorporada en juegos serios", en *Juegos serios: mecanismos y efectos*, U. Ritterfeld, MJ Cody y P. Vorderer, Eds., Págs. 322–343, Routledge, Nueva York, NY, EE. UU., 2009.

[12] Massa, S. y Kühn F. *Analíticas de aprendizaje para Serious Games*. XX WICC.

[13] Ferguson, R. y Shun, S. *Social Learning Analytics*. *Journal of Educational Technology and Society*. July 2012.

[14] Massa, S. M., Moro, L., Bacino, G., Pirro, A., Evans, F., Hinojal, H., Spinelli, A., Zapiraín, E., Rico, C., Kühn, F., Lanzillota, F. *El proceso de desarrollo de serious games. Modelos, herramientas y analíticas de aprendizaje*. WICC 2020.

[15] *The Impact of Choice and Feedback on Learning, Motivation, and Performance in an Educational Video Game*

[16] Hattie, J., Timperley, H. *The Power of Feedback*. *Review of Educational Research*, Vol. 77, No. 1 (Mar., 2007), pp. 81-112. American Educational Research Association.

[17] BlackyWilliam. *Developing the theory of formative assessment*. *Educational Assessment Evaluation and Accountability*. Febrero 1998. DOI: 10.1007/s11092-008-9068-5

[18] Alegre, A. *Academic self-efficacy, self-regulated learning and academic performance in first-year university students*. *Propósitos y Representaciones*, 2(1), 79-120. doi:

<http://dx.doi.org/10.20511/pyr2014.v2n1.54>

[19] Gaycho-Rodríguez, T. *Retroalimentación del desempeño frente al retroalimentación del resultado: un estudio inicial de sus efectos sobre el aprendizaje*. *Educationis Momentum*. Vol 3. 2017. ISSN 2414-1364.

[20] Alvarado, M. *Retroalimentación en educación en línea: una estrategia para la construcción del conocimiento*. RIED. *Revista Iberoamericana de Educación a Distancia*, vol. 17, núm. 2, 2014, pp. 59-73 Asociación Iberoamericana de Educación Superior a Distancia Madrid, Organismo Internacional.

Investigaciones y experiencias en el área de Interacción Persona Ordenador y Educación

Sanz Cecilia^{1,4} , Gorga Gladys¹ , Artola Verónica^{1,3} , Salazar Mesía Natalí¹ , Iglesias Luciano¹ , Archuby Federico^{1,2} , Nordio Mauricio^{1,4} , Buffarini Abril¹ , Ibañez Bárbara^{1,2}, Astudillo Gustavo⁵ 
Baldassarri Sandra⁶ 

¹Instituto de Investigación en Informática LIDI (III-LIDI). Centro Asociado CIC.
Facultad de Informática – Universidad Nacional de La Plata

²Becaria UNLP

³Becaria Doctoral CONICET

⁴ Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

⁵ Facultad de Ciencias Exactas y Naturales - Universidad Nacional de La Pampa

⁶GIGA AffectiveLab, Universidad of Zaragoza, España

{csanz, vartola, nsalazar, farchuby, mnordio}@lidi.info.unlp.edu.ar,

li@info.unlp.edu.ar, abrilbuffarini@gmail.com, astudillo.gustavo@gmail.com, sandra@unizar.es

RESUMEN

En este trabajo se presentan las investigaciones y experiencias que se vienen desarrollando en el marco de un sub-proyecto del Instituto de Investigación en Informática LIDI. En particular, se hace énfasis en los resultados alcanzados en el año 2021. El foco del trabajo se relaciona con modelos de interacción persona ordenador y su integración a contextos educativos. Específicamente, se aborda el diseño y desarrollo de juegos que integran diferentes modelos de interacción, cuyos objetivos van más allá del entretenimiento, entamándose con objetivos educativos. Al mismo tiempo, se diseñan experiencias educativas en entornos interactivos, que combinan el espacio físico y el virtual. También, se aborda el estudio de técnicas y herramientas vinculadas a la Computación Afectiva que pueden enriquecer procesos educativos mediados por tecnologías digitales. Se detallan aquí los avances realizados en estas temáticas. Como parte del proyecto, se participa en la formación de recursos humanos en el área, a través de tesis de postgrado, trabajos finales de grado, becarios de investigación y proyectos de innovación con alumnos.

Palabras clave: interacción persona-ordenador, interacción tangible, realidad aumentada, realidad virtual, entornos inmersivos, computación afectiva, escenarios educativos, juegos serios

CONTEXTO

Este trabajo se enmarca en el sub-proyecto titulado “Metodologías y herramientas para la apropiación de tecnologías digitales en escenarios educativos híbridos”, y forma parte de un proyecto más general titulado: “Metodologías, técnicas y herramientas de Ingeniería de Software en escenarios híbridos. Mejora de proceso” (período 2018-2021), en el que se estudian y diseñan

metodologías y herramientas de la Ingeniería de Software para escenarios híbridos que entran diferentes entornos, dispositivos, formas de acceso y de interacción. Se trata de un proyecto del Instituto de Investigación en Informática LIDI, de la Facultad de Informática de la Universidad Nacional de La Plata, acreditado por el Ministerio de Educación de la Nación.

1. INTRODUCCION

Las Interfaces de Usuario Tangibles (TUIs: Tangible User Interfaces) y la Interacción Tangible (IT), son conceptos que han ido ganando interés en el campo del área de Interacción Persona Ordenador (IPO) (Rodić & Granić, 2021).

Estas líneas de investigación se basan en la tangibilidad y la interacción con todo el cuerpo, integrando recursos computacionales diversos así como también diferentes formas de datos físicos y materiales (Hornecker y Buur, 2006). La IT combina el mundo real y el digital, dando la oportunidad de crear experiencias interactivas. Como consecuencia de estas posibilidades, en la actualidad se está trabajando en narrativas tangibles que permiten también abordar objetivos educativos. Éstas reflejan una realidad física del mundo, una historia o ambas (Echeverri & Wei, 2021). Las narrativas tangibles permiten interacciones naturales, ubicando en el centro de la experiencia a la persona, quien explora el espacio físico e interactúa con objetos aumentados digitalmente. Estos aspectos son de importancia para el aprendizaje (Marshall, 2007). El trabajo presentado por Rodić y Granić (2021), a través de una revisión sistemática de literatura, da cuenta de cómo se vienen desarrollando diversas experiencias educativas usando interacción tangible, y en particular, narrativas tangibles. Estos autores indican que las TUIs se constituyen en una ayuda para procesos de aprendizaje, ya que integran la dimensión háptica y el uso de un espacio compartido que combina lo digital con lo físico, y genera motivación y,

a su vez, da la posibilidad de crear nuevos materiales educativos (*digital manipulatives u objetos activos*) (Zuckerman, Arida, & Resnick, 2005). Los objetos activos permiten el pasaje de representaciones físicas a virtuales que puedan aportar al razonamiento y la comprensión. En el trabajo que aquí se presenta, se viene trabajando en el diseño y desarrollo de juegos que utilizan narrativas tangibles, con interfaces que involucran una mesa interactiva (*tabletop*), objetos activos (con capacidad de acción a partir de diferentes actuadores y sensores) (Alvarado, Sanz & Baldassarri, 2021), y proyecciones (Iglesias & Sanz, 2020), de manera tal de ofrecer una experiencia educativa innovadora.

Al mismo tiempo, se aborda el diseño de juegos educativos que utilizan realidad aumentada o realidad virtual como modelo de interacción. En el caso de la realidad aumentada, en el marco del proyecto, se está diseñando aplicaciones móviles con dinámica de juegos educativos. El uso de dispositivos móviles facilita el aumento de elementos físicos con escenas digitales, y posibilita el desarrollo de actividades interesantes con estudiantes. A partir de diversos estudios que aplican esta tecnología en el contexto universitario se puede observar que promueve la motivación, el aprendizaje activo, y constructivo (Laurens Arredondo, 2022; Romano, Sanz y Gorga, 2021). También se vienen utilizando técnicas de proyecciones para lograr aplicaciones de realidad aumentada. Esto se está aplicando en experiencias educativas que buscan enriquecer el contexto real con información digital (De Russis, 2015).

En cuanto a la realidad virtual (RV), ésta permite simular la presencia física de personas y objetos representados de manera gráfica en tres dimensiones, y genera así experiencias sensoriales realistas, posibilitando la sensación de inmersión, presencia e interactividad (Anopas, & Wongsawat, 2014), propias de esta tecnología (Menjivar Valencia et al., 2021; Chirinos, Sanz, & Dapoto, 2020). Durante 2020 y 2021, se han desarrollado diversos juegos de RV que se pueden jugar con gafas y dispositivos móviles para ofrecer una experiencia inmersiva, y también se han exportado para PC para dar más posibilidades de acceso a los estudiantes interesados.

Finalmente, como parte de las líneas de investigación y experiencias que aquí se presentan, se está trabajando en el área de Computación Afectiva, para enriquecer procesos educativos en los que intervienen tecnologías digitales (Astudillo, Sanz & Baldassarri, 2020). Picard (1999) en su trabajo sobre Computación Afectiva, menciona esta disciplina como vinculada al área de Inteligencia Artificial, y afirma que se orienta a desarrollar métodos computacionales vinculados a reconocer emociones humanas y generar emociones sintéticas. En este sentido, en el proyecto que da contexto a este trabajo, se vienen desarrollando tesis doctorales basadas a esta temática y su aplicación en escenarios educativos.

Las líneas de investigación aquí presentadas guardan relación directa con las temáticas de la Maestría y Especialización en Tecnología Informática Aplicada en Educación de la misma Facultad, por lo que se desarrollan tesis y trabajos finales en estos temas. También se cuenta con una agenda de cooperación a nivel nacional e internacional, que permite profundizar el trabajo que se lleva adelante en el III-LIDI.

2. LINEAS DE INVESTIGACION / DESARROLLO

Se mencionan aquí las principales líneas de investigación y desarrollo abordadas en el marco del proyecto:

- Juegos educativos basados en realidad aumentada, interacción tangible y realidad virtual. Juegos pervasivos.
- Metodologías y herramientas para el diseño de aplicaciones educativas basadas en estos paradigmas.
- Interacción Tangible. Marcos que fundamentan este paradigma, relación entre objetos físicos y digitales, aplicaciones en educación.
- Objetos Pasivos y Activos para IT. Tipos de *feedback* en objetos activos.
- Realidad Aumentada para juegos educativos.
- Entornos inmersivos e interactivos basados en realidad virtual. *Embodied Interaction*
- Sistemas educativos adaptativos y sistemas recomendadores para recursos educativos.
- Computación Afectiva en entornos digitales para el escenario educativo.

Estas líneas se sostienen a partir del trabajo conjunto de docentes investigadores, becarios, tesistas, pasantes y estudiantes que participan.

3. RESULTADOS OBTENIDOS

En esta sección se presentan los resultados alcanzados en el sub-proyecto fundamentalmente durante el 2021 e inicios del 2022.

En 2021 se profundizó una revisión de literatura sobre experiencias educativas de IT, al mismo tiempo se avanzó en el análisis de los resultados de la tesis doctoral de uno de los miembros del proyecto (Artola, Sanz & Pesado, 2020; Artola, Pesado & Sanz, 2019). Durante 2021, se elaboró un artículo que está siendo evaluado (Artola, Sanz & Baldassarri, 2022). Se abordaron pruebas con estudiantes y otros usuarios del juego serio Albores, que integra interfaces tangibles con objetos activos, y realidad aumentada, creando un ambiente combinado físico-virtual. Albores fue presentado como demo educativa en el marco del Congreso TEyET 2021 (Bigurrarena et al., 2021). Este juego se orienta a dar a conocer figuras innovadoras de la historia de la Informática, e integra el uso de una mesa interactiva, donde se desbloquean mini-juegos a medida que se avanza en las categorías y niveles que se proponen. Las pruebas realizadas han llevado a mejorar alguna de las tecnologías subyacente para dar más robustez en la

interacción. Aquí se puede visualizar una demo de este juego (<https://youtu.be/cU009RmpW7g>). También se llevan adelante tareas técnicas relacionadas con la construcción de mesas interactivas (Nordio & Sanz, 2021). Actualmente, se está terminando de armar una tercera mesa (VisionAR III).

En el marco de un proyecto de innovación con estudiantes, se llevó adelante un juego denominado Escapados, el cual integró un equipo interdisciplinario en el que intervinieron también estudiantes de la Especialización y Maestría en Tecnología Informática Aplicada en Educación. Este juego se orienta a reforzar aprendizajes del área de Química y Matemática, a través de un entorno 3D, y el uso de interacción tangible con objetos del entorno físico que deben ser reconocidos ante la cámara web del jugador, para ser incorporados como parte de la historia de Escapados y poder resolver desafíos. Para ello fue necesario utilizar técnicas de reconocimiento basadas en visión por computadora. El juego aún está en desarrollo, y fue presentado en las Jornadas de Ciencia y Tecnología 2021 de la Facultad de Informática, UNLP (Zeballos et al., 2021).

En cuanto a las investigaciones y experiencias con RV, durante 2021 se avanzó en los temas relacionados a la tesis de maestría de Chirinos (2020). Además, se continuó articulando con el proyecto de extensión de la UNLP, titulado Huellas patrimoniales (Rucci & Sanz, 2021). Este proyecto se orienta a dar a conocer y acercar el patrimonio argentino a poblaciones (en especial niñas/niños) con algún tipo de vulnerabilidad social. En 2021 se desarrollaron dos ejes más del juego de realidad virtual: HuVi Parque Nacional Iguazú. Se trabajó en metodologías de diseño de juegos serios de RV, previo a abordar la implementación (Chirinos & Sanz, 2021; Archuvy, Sanz, & Manresa-Yee, 2020; Chirinos, Sanz, Rucci, Comparato, Gonzalez, Dapoto, 2020). También se diseñó en forma interdisciplinaria la versión de HuVi Ischigualasto y Talampaya. Esto involucró un trabajo entre miembros del proyecto Huellas Patrimoniales de la Licenciatura en Turismo, otros de la Facultad de Artes, y de la Facultad de Informática que cooperaron en el guionado, la producción multimedia y el diseño y desarrollo de la app de RV (Mazza, Rucci & Sanz, 2021). Se ofrecieron charlas y actividades en torno a estas aplicaciones. Además, se trabaja en dos tesis doctorales vinculadas al diseño y desarrollo de juegos serios, y en particular, de realidad virtual.

En 2021 y lo que va de 2022, se puso en marcha la difusión de Innovática, otro juego de RV desarrollado en el marco de un trabajo final de licenciatura, en el contexto de este sub-proyecto de investigación del LIDI (Mazza, Sanz & Artola, 2020). Este juego fue evaluado considerando su usabilidad, además, se está aplicando en talleres en el Centro de Innovación y Transferencia Tecnológica (CIyTT) de la Facultad de Informática con estudiantes y docentes.

En relación a los temas de RA, se avanzó en una tesis de maestría donde se estudia la incidencia de la integración de actividades con realidad aumentada en el rendimiento académico (Romano, Sanz & Gorga, 2021). La tesis fue

presentada a fines de 2021 y está esperando su evaluación. En el marco de esta tesis se diseñó una actividad educativa basada en una guía de práctica en temas de programación, en la que se integró la realidad aumentada. El aumento presenta a un robot con guías pedagógico-afectivas para acompañar las tareas de los estudiantes. También se está desarrollando una tesis de maestría en la que se propone una herramienta para que los docentes desarrollen actividades educativas basadas en RA a partir de geolocalización (Salazar, Sanz, & Gorga, 2019, a y b). Finalmente, en 2021 se contó con una beca de iniciación a las vocaciones científicas, de la UNLP, a partir de ésta, se trabajó en el diseño de EmpoderAR (Ibañez & Sanz, 2021). Esta aplicación móvil, siguió estrategias de diseño basadas en las investigaciones previas de este sub-proyecto. En particular, se basó en la experiencia previa de diseño de Ruta Darwin (Lizarralde et al., 2019). EmpoderAR permite a través de una línea de tiempo sobre figuras destacadas de la Informática, aumentar la historia de algunas de las mujeres que han sido innovadoras en esta disciplina. Se trabaja con una dinámica lúdica, que a través de trivias, que posibilitan que el jugador puede ganar trofeos que luego aparecerán en la cueva de la sabiduría, la cual forma parte de la historia de EmpoderAR.

En relación a los temas de Computación Afectiva, se dirigen dos tesis doctorales en las que se trabaja en la recomendación de recursos (música, y videos educativos) considerando las emociones (Astudillo et al., 2020; Ospitia-Medina et al., 2020). En 2021 se ha investigado sobre los sesgos en los sistemas recomendadores de música (Ospitia-Medina et al., 2021, 2020). Al mismo tiempo, se investigaron metadatos para caracterizar videos educativos con emociones (Astudillo, Sanz & Baldassarri, 2021a), y sobre bases de datos emocionales, a partir de una revisión sistemática de literatura (Astudillo, Sanz & Baldassarri, 2021b). Finalmente, se está participando en estas temáticas a través de otros proyectos con universidades del país (Sanz et al., 2021) y del exterior (Baldassarri, Sanz, Coma, Aguelo-Arguis, & Alvarez, 2019).

En cuanto a los proyectos vinculados y los acuerdos de cooperación, el III-LIDI participa en los siguientes:

- Se cuenta con un acuerdo de colaboración en estos temas con la Universidad de Zaragoza y la Universidad de Islas Baleares para cooperar en estas temáticas.
- Se ha obtenido financiamiento en el marco de una convocatoria K107 de Erasmus para estancias de viaje entre la Universidad de Zaragoza y la Universidad Nacional de La Plata.
- Se participa en el proyecto “*Pervasive Gaming Experiences For @ll* (Pergamex)”, en particular en el subproyecto (RTI2018-096986-B-C31) de la UZ.
- Además, la Dra. Sanz es miembro colaborador del grupo de Investigación en Interfaces Avanzadas (AffectiveLab).
- Se participa en la Red constituida por universidades de Iberoamérica en el marco del programa "Pablo Neruda" dentro del Espacio Iberoamericano del

Conocimiento (EIC) y de la Organización de los Estados Iberoamericanos (OEI) orientada a la movilidad de estudiantes y docentes de doctorado.

- Se participa en la RedAUTI: Red temática en Aplicaciones y Usabilidad de la Televisión Digital Interactiva. En este ámbito se analizan materiales educativos para la TVDI.
- Se dirige un proyecto con la Universidad Nacional de Río Negro en relación a estas temáticas y se participa asesorando un proyecto de la Universidad Nacional de Santiago del Estero.

4. FORMACION DE RECURSOS HUMANOS

En este proyecto se participa en la formación de recursos humanos a través de la dirección de tesis de doctorado, maestría, trabajos finales de especialización y tesinas de grado; también con becas de diferentes organismos de ciencia y técnica. En 2021, se ha finalizado 1 tesis de Maestría dirigida en relación a los temas de RA y que espera su evaluación. Se ha finalizado una beca de iniciación a las vocaciones científicas de UNLP, con presentación e informe final de actividades. Se trabajó en la formación de un personal de apoyo de la CICPBA vinculado a los temas de Interacción Tangible. Además se ha desarrollado un proyecto de innovación con alumnos, con 3 alumnos participantes de Ingeniería en Computación, y 2 de Especialización y Maestría en TIAE, respectivamente.

5. BIBLIOGRAFIA

- Alvarado, M. C., Sanz, C., & Baldassarri, S. (2021). Desarrollo de una experiencia educativa con juguetes activos en una mesa de interacción tangible. *Revista de la Asociación Interacción Persona Ordenador (AIPO)*, 2(2), 74-84.
- Anopas, D., & Wongsawat, Y. (2014). Virtual reality game for memory skills enhancement based on QEEG. *Proceedings of the 7th 2014 Biomedical Engineering International Conference*, pp. 1–5
- Archuby, F., Sanz, C. & Manresa-Yee, C. (2020). Metodologías de diseño y desarrollo para la creación de juegos serios digitales. Tesis de maestría finalizada. Sep. 2020. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/111123>
- Artola, V., Sanz, C. & Baldassarri, S. (2022). A Novel Tangible Interaction Authoring Tool for Creating Educational Activities: Analysis of its Acceptance by Educators. *TLT-2022-01-0025*. Enviado para su evaluación en enero de 2022.
- Artola, V., Sanz, C. & Pesado, P. (2020). Tesis Doctoral. Interacción tangible en escenarios educativos. Diseño de una herramienta de autor para la creación de Actividades educativas basadas en interacción tangible. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/98135>
- Artola, V.; Pesado, P. & Sanz, C (2019). EDIT: una herramienta de autor para la creación de actividades educativas basadas en interacción tangible. *Actas del XIV Congreso sobre Tecnología en Educación & Educación en Tecnología (TE&ET 2019)*, ISBN: 978-987-733-196-7, págs. 115-124. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/90735>
- Astudillo, G. J., Sanz, C. V., & Baldassarri Santalucía, S. (2021a). Revisión sistemática sobre la meta- anotación de videos educativos con emociones. In *XVI Congreso de Tecnología en Educación & Educación en Tecnología-TE&ET 2021 (La Plata, 10 y 11 de junio de 2021)*.
- Astudillo, G., Sanz, C. & Baldassarri, S. (2021b). Análisis del etiquetado emocional de videos educativos. *JAUTI* 2021. <https://link.springer.com/conference/jatui>
- Astudillo, G.; Sanz, C. & Baldassarri, S. (2020). Definición de un modelo de relaciones entre tipos de videos educativos, perfiles de usuarios y emociones. Propuesta de tesis de doctorado aprobada en 2020.
- Baldassarri, S., Sanz, C. Coma, T., Aguelo, A. & Alvarez, P. (2019) Involving students in the generation of automatic assessment tests. *12th annual International Conference of Education, Research and Innovation. ICERI2019*, isbn 978-84-09-14755-7, issn 2340-1095, doi: 10.21125/iceri.2019.1687, España, 7105-7113
- Bigurrarena, N., Ballardini, E., Artola, V., Buffarini, A., Nordio, M., & Sanz, C. V. (2021). Albores: un juego basado en interacción tangible para conocer figuras destacadas de la historia de la Informática. In *XVI Congreso de Tecnología en Educación & Educación en Tecnología-TE&ET 2021 (La Plata, 10 y 11 de junio de 2021)*.
- Chirinos, Y., & Sanz, C. (2021). Guión ejes 2 y 3 HuVi Parque Nacional Iguazú.
- Chirinos, Y., Sanz, C. & Dapoto, S. (2020) La realidad virtual como mediadora de aprendizajes. Desarrollo de una aplicación móvil de realidad virtual orientada a niños. Tesis de Maestría en Tecnología Informática Aplicada en Educación finalizada en Diciembre de 2020. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/111879>
- Chirinos Delfino, Y., Sanz, C. V., Rucci, A. C., Comparato, G. J., Gonzalez, G., & Dapoto, S. H. (2020). HUVI: una aplicación de realidad virtual para acercar el patrimonio argentino. In *XV Congreso Nacional de Tecnología en Educación y Educación en Tecnología (TE&ET 2020)*.
- De Russis, L. (2015). *Interacting with Smart Environments: Users, Interfaces, and Devices*. IOS Press. Vol7, Nro.1, ISSN: 1876-1364. *J. Ambient Intell. Smart Environ*, pp 115-116.
- Echeverri, D. & Wei, H. (2021). Designing Physical Artifacts for Tangible Narratives: Lessons Learned from Letters to José In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '21)*. Association for Computing Machinery, New York, NY, USA, Article 94, 1–12. DOI: <https://doi.org/10.1145/3430524.3446070>

- Hornecker, E. & Buur, J. (2006). Getting a Grip on Tangible Interaction: A Framework on Physical Space and Social Interaction. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI 2006, ACM, Montreal, Quebec, Canada, 437–446.
- Ibañez, B.R. & Sanz, C. (2021). Informe final de beca. Entregado en octubre de 2021.
- Iglesias, L. & Sanz, C. (2020). Aplicación de inmersión en Murales con Kinect. Informe de Tareas.
- Laurens-Arredondo, L. (2022). Mobile augmented reality adapted to the ARCS model of motivation: a case study during the COVID-19 pandemic. *Educ Inf Technol*. <https://doi.org/10.1007/s10639-022-10933-9>
- Lizarralde, A.; Sanz, C.; Gorga, G.; Buffarini, A.; Beltrán, E. & Kraselsky, R. (2019). Ruta Darwin: un juego con realidad aumentada para conocer las experiencias de Charles Darwin en su travesía a bordo del Beagle. *TE&ET 2019*. ISBN: 978-987-733-196-7. Pp. 262-264.
- Sanz, C. V., Lovos, E., Goin, M., Ricca, M. V., Molina, C., Gil, E., & Gastaminza, M. (2021). Juegos serios y realidad aumentada. In XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021, Chilecito, La Rioja).
- Marshall, P. (2007). Do tangible interfaces enhance learning? En Proceedings of the 1st international conference on tangible and embedded interaction (pp. 163–170). New York, USA: ACM. <http://doi.acm.org/10.1145/1226969.1227004> doi: 10.1145/1226969.1227004
- Mazza, M., Rucci, A.C., & Sanz, C. (2021). Juego serio con realidad virtual HuVi Ischigualasto y Talampaya. Desarrollada con el equipo del proyecto Huellas Patrimoniales. Disponible en: https://play.google.com/store/apps/details?id=com.LIDI.HuViIschigualastoTalampaya&hl=es_AR&gl=US
- Mazza, M., Sanz, C., & Artola, V. (2020). Juego serio con realidad virtual para jóvenes orientado a conocer hitos de la historia de la Informática. Tesina de grado finalizada. Disponible en:
- Menjívar Valencia, E., Sánchez Rivas, E., Ruiz Palmero, J., & Linde Valenzuela, T. (2021). Revisión de la producción científica sobre la Realidad Virtual entre 2016 y 2020 a través de Scopus y WOS. *EDMETIC*, 10(2), 26-55. <https://doi.org/10.21071/edmetic.v10i2.13422>
- Nordio, M. & Sanz, C. (2021). Informe técnico de tareas de personal de apoyo CIC. Período: octubre 2020 a octubre 2021.
- Ospitia-Medina, Y., Baldassarri, S., Sanz, C., & Beltrán, J. (2021). Music Recommender Systems: A Review Centered on Biases. Aprobado para publicar en “Advances in Speech and Music Technology: Computational Aspects and Applications” of the Springer book series “Signal and Communication Technology”. Esperando su publicación.
- Ospitia-Medina, Y., Baldassarri, S., Sanz, C., Beltrán, J. R., & Olivas, J. A. (2020). Fuzzy Approach for Emotion Recognition in Music. In 2020 IEEE Congreso Biental de Argentina (ARGENCON) (pp. 1-7). IEEE.
- Picard, R. (1999). Affective Computing for HC, en Proceedings of HCI International on Human-Computer Interaction: Ergonomics and User Interfaces-Volume I - Volume I, Hillsdale, NJ, USA, 1999, pp. 829-833.
- Rodić, L. D., & Granić, A. (2021). Tangible interfaces in early years’ education: a systematic review. *Personal and Ubiquitous Computing*. doi:10.1007/s00779-021-01556-x
- Romano, L., Sanz, C. V., & Gorga, G. M. (2021). Realidad aumentada y su vinculación con el rendimiento académico. In XVI Congreso de Tecnología en Educación & Educación en Tecnología-TE&ET 2021 (La Plata, 10 y 11 de junio de 2021).
- Rucci, A.C. & Sanz, C. (2021). Informe de tareas proyecto de Extensión: Viajes inclusivos y nuevas experiencias sensoriales por medio de realidad virtual.
- Salazar, N.; Sanz, C. & Gorga, G. (2019 a). Análisis comparativo de librerías de realidad aumentada. Sus posibilidades para la creación de actividades educativas. *Especialización en TIAE*.
- Salazar Mesía, N.; Sanz, C. & Gorga, G. (2019 b). Diseño de plantillas para la creación de actividades educativas con Realidad Aumentada en AuthorAR. XXV Congreso Argentino de Ciencias de la Computación. Octubre, Córdoba. ISBN: 978-987-688-377-1.
- Salazar Mesía, N.; Sanz, C. & Gorga, G. (2019 c). Posibilidades de las librerías de Realidad Aumentada en el desarrollo de actividades educativas. XIV Congreso Tecnología en Educación y Educación en Tecnología (TEYET 2019). San Luis.
- Zeballos, M., Lombardo, S., Fanelli, V., Artola, V., Archuby, F., Ferreyra, P., Gubaro, M., Sanz, C. (2021). Escapados. Presentación de la Jornadas de CyT 2021, FI, UNLP. Disponible en: <https://youtu.be/ffqFTiPult4>
- Zuckerman, O., Arida, S., y Resnick, M. (2005). Extending tangible interfaces for education: Digital montessori-inspired manipulatives. En Proceedings of the sigchi conference on humanfactors in computing systems, pp. 859–868).ACM. <http://doi.acm.org/10.1145/1054972.1055093>

Realidad Aumentada en diálogos de estudiantes colaborativos

Luis Gerez Martinez, Rosanna Costaguta y María de los Ángeles Menini
Instituto de Investigación en Informática y Sistemas de Información (IIISI)
Facultad de Ciencias Exactas y Tecnologías (FCEyT)
Universidad Nacional de Santiago del Estero (UNSE)
luis.gerez@gmail.com; {rosanna, marameni}@unse.edu.ar

RESUMEN

Los beneficios que los entornos de Aprendizaje Colaborativo Soportado por Computadora aportan a los procesos de enseñanza y de aprendizaje están ampliamente investigados. Sin embargo, la inclusión tecnologías actuales en estos entornos, como es la realidad aumentada, abrió nuevos caminos de investigación.

Docentes y estudiantes de la Universidad Nacional de Santiago del Estero disponen desde hace varios años de la aplicación COLLAB. Esta aplicación fue especialmente creada para dar soporte a las actividades de diálogo en pequeños grupos de estudiantes colaborativos.

La presente línea de investigación se enfoca en el estudio de la realidad aumentada aplicada en contextos colaborativos de educación. En particular, se quiere incorporar nuevas funcionalidades a la aplicación COLLAB para posibilitar la gestión de recursos aumentados por parte de docentes y estudiantes. Para esto, se desarrollará especialmente un módulo que luego será acoplado a COLLAB. La experimentación con grupos de estudiantes reales, que usarán la aplicación durante sesiones colaborativas especialmente diseñadas, permitirá recopilar los datos necesarios para efectuar la validación correspondiente.

Palabras clave: *Realidad Aumentada, Aprendizaje Colaborativo Soportado por Computadora, Tecnologías web 2.0 y móviles.*

CONTEXTO

En este artículo se presenta una de las líneas de investigación del proyecto “*Desarrollo de aplicaciones para colaboración en e-learning*”, acreditado y financiado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Santiago del Estero (SECyT - UNSE) para el período 2022-2025. El proyecto responde a una de las líneas prioritarias de investigación del IIISI, vinculada con la creación y uso de tecnologías informáticas destinadas a mejorar los procesos de enseñanza y de aprendizaje. Por otro lado, el proyecto continúa una línea de investigación iniciada en 2010-2011 por el proyecto 23/C089 “*Fundamentos Conceptuales y Soportes Tecnológicos de la Informática Educativa*”, continuada en 2012-2016 por el proyecto 23/C097 “*Sistemas de Información Web Basados en Agentes para Promover el Aprendizaje Colaborativo Soportado por Computadora*”, y en 2017-2021 por el proyecto 23/C138 “*Mejorando escenarios de aprendizaje colaborativo soportado por computadora*”. Todos proyectos acreditados y financiados por SECyT – UNSE.

1. INTRODUCCIÓN

En los últimos años, de la mano de internet y del avance tecnológico, los contextos pedagógicos y sus propias prácticas han ido evolucionando. Estas transformaciones llevaron a definir nuevos espacios formativos y a implementar en ellos nuevas tecnologías que permiten generar nuevos recursos utilizables en los procesos de enseñanza y de aprendizaje. Este es el caso de la denominada Realidad Aumentada (RA).

Azuma [1] definió a la RA como sistemas que presentan las siguientes características: combinan lo real con lo virtual, existe interacción en tiempo real, y esa interacción es tridimensional.

Por otra parte, para Cubillo *et al.* [2] la RA es todo sistema interactivo que tiene como entrada la información del mundo real, a la que superpone nueva información digital en tiempo real, pudiendo esa información ser imágenes, objetos tridimensionales, textos, videos, etc. Con lo que la realidad se presenta al usuario aumentada a través del uso de diferentes tecnologías.

Pence [3] establece la existencia de dos taxonomías de aplicación de RA en términos de desarrollo: la RA basada en marcadores y la RA sin marcadores.

Un marcador es una etiqueta o clave que permite activar un objeto virtual posicionándolo en el mundo real de los usuarios. La RA basada en marcadores requiere para operar de cinco elementos: un dispositivo informático (por ejemplo, una computadora portátil o de escritorio), un software para RA, etiquetas o marcadores, cámara web, y un dispositivo de visualización (por ejemplo, una pantalla de computadora o un proyector). Al detectar el marcador a través de la cámara web, el objeto virtual se puede mostrar en la pantalla de la computadora.

Por el contrario, la RA sin marcadores se basa en cualquier parte del entorno real como objetivo para proyectar un objeto virtual. Opera incorporando un sistema de posicionamiento global (GPS) que identifica coordenadas o ubicaciones. Luego, el GPS entrega la información generada por computadora a esas coordenadas [4].

Generalmente todos los elementos enunciados en los párrafos previos están disponibles en los ambientes virtuales de aprendizaje, lo cual facilita la incorporación de la RA en ellos.

Estudios recientes han reportado que la inclusión de nuevos elementos tecnológicos con fines educativos (como la RA), aumenta el

interés por aprender y la motivación de los estudiantes [5]. Adicionalmente, la utilización de aplicaciones colaborativas con RA fomenta el trabajo en equipo, lo que resulta ventajoso pues así los estudiantes comparten conocimiento, dudas, u opiniones, y logran un mejor nivel cognitivo al que obtienen trabajando individualmente [6].

Actualmente, los docentes y estudiantes de la Universidad Nacional de Santiago del Estero disponen de una aplicación web especialmente creada para gestionar y soportar experiencias de aprendizaje colaborativo soportado por computadora basadas en diálogo (síncrono y asíncrono). Esta aplicación, denominada COLLAB [7], fue desarrollada años atrás por otra de las líneas de trabajo del proyecto de investigación al que pertenece la presente propuesta. COLLAB funciona en los servidores de la universidad para facilitar el acceso de la comunidad universitaria.

El objetivo de esta investigación es ampliar las funcionalidades de COLLAB para que los estudiantes y docentes que lo usan puedan valerse de recursos aumentados en sus actividades de enseñanza y de aprendizaje.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO/OBJETIVOS

Esta línea de trabajo se inserta en el área del aprendizaje colaborativo soportado por computadora, buscando promover la inclusión de nuevas tecnologías informáticas, para mejorar los procesos de enseñanza y de aprendizaje.

Puntualmente, se quiere acrecentar las posibilidades de uso de tecnologías en la aplicación COLLAB, para que los estudiantes que dialogan distribuidos en pequeños grupos puedan disponer de material de estudio aumentado (artículos en formato PDF, videos o imágenes). Por lo expuesto, se establecieron los siguientes objetivos para esta línea de trabajo:

- Desarrollar un módulo en entorno de RA para ser incluido en la aplicación COLLAB.

- Lograr que sea un módulo híbrido, es decir que funcione tanto para entornos WEB2.0, con sistemas operativos Windows y Linux, como para dispositivos móviles, con sistemas operativos Android e Iphone.
- Incluir una función que permita a los profesores cargar los recursos aumentados que quedarán disponibles para los grupos de estudiantes.
- Permitir a los estudiantes acceder a los recursos aumentados de manera personalizada para luego trabajar de manera colaborativa con sus compañeros de grupo.

Para alcanzar estos objetivos, primero se realizará la búsqueda y el análisis de bibliografía vinculada con el aprendizaje colaborativo soportado por computadora y con el desarrollo de aplicaciones con realidad aumentada en contextos educativos. Luego se llevará a cabo el diseño del nuevo módulo, y su implementación en la aplicación COLLAB. Para finalizar, se validará su correcto funcionamiento mediante su uso por parte de estudiantes y docentes reales.

3. RESULTADOS ESPERADOS

Esta investigación pretende favorecer el desarrollo de conocimiento científico-tecnológico de relevancia para el desarrollo de sistemas de información web que hagan uso de recursos aumentados, adaptables a dispositivos móviles, en el área del Aprendizaje Colaborativo Soportado por Computadora.

Algunos de los resultados a los que se espera arribar al concretar este trabajo son:

- Estado del arte referido a la inclusión de RA en contextos educativos.
- Una ampliación de las actuales funcionalidades de la aplicación COLLAB, a través de un nuevo módulo que permita gestionar recursos aumentados para uso en experiencias universitarias de aprendizaje colaborativo soportado por computadora.
- Resultados experimentales que permitan validar la propuesta.
- Difusión de los resultados alcanzados mediante presentaciones en congresos y publicación de artículos en revistas.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo de esta línea de investigación está compuesto por un docente investigador formado, y un docente investigador en formación. El equipo también cuenta con un integrante estudiante que está desarrollando su tesis de grado para obtener el título de Licenciado en Sistemas de Información en el marco de esta investigación.

5. REFERENCIAS

- [1] Azuma, R. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environment*, Vol. 6 (4), pp. 355–385. DOI:<https://doi.org/10.1162/pres.1997.6.4.355>
- [2] Cubillo, J., Martín, S., Castro, M. & Colmenares, A. (2014). Recursos digitales autónomos mediante realidad aumentada. *Revista Iberoamericana de Educación a Distancia*, Vol. 17(2), pp. 241-274. DOI:<https://doi.org/10.5944/ried.17.2.12686>
- [3] Pence, H. (2011). Smartphones, Smart Objects, and Augmented Reality. *The Reference Librarian*, Vol. 52, pp. 136-145. DOI:<https://doi.org/10.1080/02763877.2011.528281>
- [4] Phon, D., Ali, M., & Halim, N. (2014). Collaborative Augmented Reality in Education: A Review. In *Proceedings of the Inter. Conference on Teaching and Learning in Computing and Engineering*, pp. 78-83, DOI:<https://doi.org/10.1109/LaTiCE.2014.23>
- [5] Barraza Castillo, R., Cruz Sánchez, V. & Vergara Villegas, O. (2015). A Pilot Study on the Use of Mobile Augmented Reality for Interactive Experimentation in Quadratic Equations. *Mathematical Problems in*

Engineering, Vol. 2015, Article ID 946034.
DOI:<https://doi.org/10.1155/2015/946034>

[6] Mendoza Morán, V., Rivera Guevara, R. & Barriga Andrade (2016). Sistemas de aprendizaje colaborativo móvil con realidad aumentada. *Revista Politécnica*, Vol. 38 (1), pp. 67-76. Editorial: Escuela Politécnica Nacional, Quito, Ecuador.

[7] Lescano, G. & Costaguta, R. (2018). COLLAB: Conflicts and Sentiments in chats. In *Proceedings of the XIX International Conference on Human Computer Interaction (Interacción 2018)*. Association for Computing Machinery (ACM), New York, USA, Article 33, pp. 1-4.
DOI:<https://doi.org/10.1145/3233824.3233864>

Adaptaciones en Aprendizaje Colaborativo Soportado por Computadora

Leandro Nicolás García, Nevelin Irene Salazar y Rosanna Costaguta
Instituto de Investigación en Informática y Sistemas de Información (IISI)
Facultad de Ciencias Exactas y Tecnologías (FCEyT)
Universidad Nacional de Santiago del Estero (UNSE)
leandrogarciasgo@gmail.com; {nsalazar, rosanna}@unse.edu.ar

RESUMEN

Dentro de un grupo de aprendizaje colaborativo cada estudiante manifiesta un comportamiento particular al desarrollar las actividades y tareas diseñadas por el profesor. Ese comportamiento no depende sólo de las necesidades que el grupo tenga en ese momento en particular, sino también, de la inclinación de cada estudiante hacia estilos propios de aprendizaje y de personalidad, entre otros.

Existen algunas investigaciones que demuestran que el estilo de aprendizaje de los estudiantes puede ser una característica a tener en cuenta para adaptar los contenidos y/o la navegación en contextos de enseñanza-aprendizaje mediados por tecnología. Sin embargo, existen otras características que podrían considerarse para ello, como, por ejemplo, los estilos de personalidad.

Docentes y estudiantes de la Universidad Nacional de Santiago del Estero disponen desde hace varios años de la aplicación COLLAB. Esta aplicación fue especialmente creada para dar soporte a las actividades de diálogo en pequeños grupos de estudiantes colaborativos.

Esta línea de trabajo propone definir adaptaciones para grupos de estudiantes colaborativos mediados por tecnología, con base en sus estilos de aprendizaje y de personalidad, a fin de maximizar sus logros de aprendizaje. La validación de tales adaptaciones será efectuada mediante sus implementaciones en la aplicación COLLAB. Posteriormente, grupos de estudiantes reales usarán la aplicación durante sesiones

colaborativas. Los datos experimentales que se recolecten serán procesados para validar esta propuesta.

Palabras clave: *Aprendizaje Colaborativo Soportado por Computadora, adaptación, estilos de aprendizaje, estilos de personalidad.*

CONTEXTO

En este artículo se presenta una de las líneas de investigación del proyecto “*Desarrollo de aplicaciones para colaboración en e-learning*”, acreditado y financiado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Santiago del Estero (SECyT - UNSE) para el período 2022-2025. El proyecto responde a una de las líneas prioritarias de investigación del IISI, vinculada con la creación y uso de tecnologías informáticas destinadas a mejorar los procesos de enseñanza y de aprendizaje. Por otro lado, el proyecto continúa una línea de investigación iniciada en 2010-2011 por el proyecto 23/C089 “*Fundamentos Conceptuales y Soportes Tecnológicos de la Informática Educativa*”, y continuada en 2012-2016 por el proyecto 23/C098 “*Sistemas de Información Web Basados en Agentes para Promover el Aprendizaje Colaborativo Soportado por Computadora*”, y en 2017-2021 por el proyecto 23/C138 “*Mejorando escenarios de aprendizaje colaborativo soportado por computadora*”. Todos los proyectos citados fueron acreditados y financiados por SECyT – UNSE.

1. INTRODUCCIÓN

El Aprendizaje Colaborativo describe la situación que se da en un grupo de estudiantes, en la que se espera que ocurran ciertas formas

de interacción entre ellos, que promuevan mecanismos de aprendizaje, pero sin garantías de que ello se produzca (Dillenbourg, 1999). Cuando los estudiantes que integran el grupo interactúan utilizando la computadora como medio de comunicación, de coordinación y de colaboración, se habla de Aprendizaje Colaborativo Soportado por Computadoras (ACSC).

La principal ventaja que se reconoce al ACSC es que los estudiantes y docentes pueden interactuar ubicados en puntos geográficos distantes, e incluso, contribuyendo en momentos diferentes en el tiempo, acorde con sus respectivas disponibilidades (Costaguta *et al.*, 2011).

Sin duda, la colaboración promueve el aprendizaje (Dillenbourg, 1999; Stahl, Koschmann & Suthers, 2006), y eso ocurre cuando los estudiantes elaboran explicaciones, comparaciones, síntesis, y son capaces de conectar diferentes ideas a través de sus interacciones (Soller, 2001; Stahl *et al.*, 2006). Sin embargo, crear grupos e instar a sus miembros a resolver una consigna de manera colaborativa, no garantiza en forma alguna que el comportamiento y el desempeño de esos grupos sean los adecuados, ni que la experiencia de enseñanza y de aprendizaje sea exitosa (Costaguta & Menini, 2014).

En la actualidad, de la mano de los recursos tecnológicos disponibles, la educación necesita proporcionar los medios y los recursos para generar nuevas formas de enseñar y de aprender. La incorporación de métodos y técnicas de adaptación en sistemas para ACSC permitirá no sólo ofrecer un aprendizaje independiente del tiempo y lugar, sino también, adaptable y personalizado a las necesidades particulares de cada estudiante y/o de cada grupo.

Por lo expuesto, en esta línea se propone investigar teórica, metodológica y experimentalmente las contribuciones que puedan realizarse para promover un ACSC exitoso mediante adaptación basada en estilos de aprendizaje y de personalización.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Según Brusilovsky (2001) la adaptación de los sistemas puede hacerse afectando su navegación y/o sus contenidos. En los sistemas de ACSC creemos que podría proporcionar a los estudiantes, por ejemplo, materiales, actividades, recorridos y asistencia de manera personalizada a sus preferencias (tanto individuales como grupales). Para ello sería necesario contar con la creación y mantenimiento de un modelo de estudiante, donde se registren los datos personales del alumno, sus progresos académicos, sus estilos de aprendizaje y de personalidad, intereses, y habilidades para resolver problemas, entre otros datos. El registro de toda esta información dará como resultado un modelo de estudiante integral, que permitirá apoyar a los alumnos a través de una personalización de todos los servicios del sistema de aprendizaje, y a los profesores ofrecer una mejor comprensión del proceso de aprendizaje de sus estudiantes (Salazar & Durán, 2016).

Los estilos de aprendizaje son un indicador de las diferencias individuales que los estudiantes poseen en la forma de percibir, representar, procesar y comprender los contenidos conceptuales enseñados por los profesores (Díaz Álvarez, 2013; Evans & Cools, 2011; Paz Penagos, 2014). Cada estudiante tiene un sistema de normas, nociones, comportamientos e ideas que le dan sentido a la práctica. Para comprender el sentido de esa práctica es necesario analizar el estilo de aprendizaje del estudiante, es decir, cómo recibe y procesa la información. Algunos antecedentes de consideración de estos estilos en sistemas de ACSC son los trabajos de Costaguta & Menini (2014), Lescano *et al.* (2016), Bourkokuou *et al.* (2016), Duque-Méndez *et al.* (2020).

Existen algunas investigaciones que vinculan a los estilos de personalidad con el aprendizaje (Bachtar *et al.*, 2017, Wu *et al.*, 2019), y algunas otras lo hacen con el desempeño académico de los estudiantes (O'Connor & Paunonen, 2007; Torres-Acosta *et al.*, 2013; Morales Rojas & Rojas Pinto,

2016; Mesurado *et al.*, 2018). Sin embargo, no se encontraron hasta el momento antecedentes que utilicen simultáneamente estilos de aprendizaje y de personalidad para realizar adaptaciones en sistemas de ACSC.

Dado lo expuesto en párrafos previos, se estableció como objetivo general de esta investigación, dotar a COLLAB de la capacidad de adaptarse a los estilos de aprendizaje y de personalidad de los estudiantes que la utilicen. Para ello se proponen las siguientes actividades:

- Estudiar las funcionalidades de la aplicación COLLAB a fin de incorporar las nuevas funcionalidades que se requieran para incorporar capacidades de adaptación y la gestión de un modelo de estudiante.
- Analizar las características de los estilos de aprendizaje y de personalidad. Se considerarán los estilos de aprendizaje definidos por Felder & Silverman (1988) y los estilos de personalidad conocidos como Big-Five (Digman, 1990).
- Diseñar el modelo de estudiante.
- Releva investigaciones que planteen vínculos entre estilos de aprendizaje y de personalidad.
- Diseñar los algoritmos de adaptación para luego programarlos e incorporarlos a COLLAB.
- Realizar experimentación para validar los desarrollos mediante el uso de COLLAB por parte de estudiantes y profesores.

Cabe aclarar que en la aplicación COLLAB los estudiantes actualmente determinan sus estilos de aprendizaje respondiendo un cuestionario compuesto por cuarenta y cuatro preguntas (Felder & Soloman, 1984) desde un enlace en la interfaz principal. De igual manera, disponen de un enlace que les permite responder otro cuestionario para descubrir sus estilos de personalidad considerando el modelo Big Five (Digman, 1990). Por lo que actualmente se dispone de esta información para cada estudiante registrado.

3. RESULTADOS ESPERADOS

Esta investigación pretende favorecer el desarrollo de conocimiento científico-tecnológico de relevancia para el desarrollo de sistemas de información web adaptativos, en el área del Aprendizaje Colaborativo Soportado por Computadora.

Algunos de los resultados a los que se espera arribar al concretar este trabajo son:

- Estado del arte referido a la inclusión de capacidades de adaptación o personalización en sistemas de e-learning.
- Una ampliación de las actuales funcionalidades de la aplicación COLLAB, dotándolo de capacidades de adaptación a los estilos de aprendizaje y de personalidad de los estudiantes.
- Resultados experimentales que permitan validar la propuesta.
- Difusión de los resultados alcanzados mediante presentaciones en congresos y publicación de artículos en revistas.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo de esta línea de investigación está compuesto por un docente investigador formado, y un docente investigador en formación. El equipo también cuenta con un integrante estudiante que está desarrollando su tesis de grado para obtener su título de Licenciados en Sistemas de Información en el marco de esta investigación.

5. REFERENCIAS

- Bachtiar, F., Sulisty, G., Cooper, E. & Katsuari, K. 2017. Affect, Personality, and Learning Styles. *In Online Reading Comprehension*. DOI: <https://doi.org/7883.10.1145/3029387.3029422>.
- Bourkougou, O., El Bachari, E. & El Alnani, M. 2016. A Personalized E-Learning Based on Recommender System. *In International Journal of Learning and Teaching*. DOI: <https://doi.org/10.18178/ijlt.2.2.99-103>.

- Brusilovsky, P. 2001. Adaptive hypermedia: User Modeling and User Adapted Interaction. *In Ten Year Anniversary Issue*. Alfred Kobsa. Amsterdam. pp. 87-110, 2001.
- Costaguta R., García P. & Amandi A. 2011. Entrenando las habilidades de colaboración de los estudiantes mediante agentes. *In Proc. IEEE Latin America Transactions*. Vol. 9(7), pp. 1118-1124. DOI: <https://doi.org/10.1109/TLA.2011.6129712>.
- Costaguta R. & Menini M. 2014. An Assistant Agent for Group Formation in CSCL based on Student Learning Styles. *In Proc. 7th Euro American Conference on Telematics and Information Systems*. Association for Computing Machinery, New York, NY, USA, Article 24, 1-4. DOI: <https://doi.org/10.1145/2590651.2590674>
- Felder, R. & Silverman, L. (1988). Learning and Teaching Styles In *Engineering Education*. *Engr. Education*, Vol. 78(7), pp. 674-681.
- Felder, R. & Soloman, V. 1984. *Index of Learning Styles*. Disponible en: <https://www.webtools.ncsu.edu/learningstyles> Último acceso: 04/03/2022.
- Díaz Álvarez, C. 2013. Mapas mentales y estilos de aprendizaje: Aportes a la enseñanza/aprendizaje en un espacio formativo en ingeniería. *Revista Educación en Ingeniería*, 8(16), pp. 45-52.
- Dillenbourg, P. 1999. Introduction: What do you mean by collaborative learning? *In Dillenbourg, P. (ed.), Collaborative learning: Cognitive and Computational Approaches*, Pergamon, Elsevier, pp. 1-19.
- Duque Méndez. N. D., Ovalle-Carranza, D., Carrillo-Ramos Á. 2020. Sistema basado en reglas para la generación personalizada de curso virtual, *TecnoLógicas*, vol. 23, no. 47, pp. 229-242, Disponible en: <https://doi.org/10.22430/22565337.1494>
- Digman, J. 1990. Personality Structure: Emergence of the Five-Factor Model. *Annual Review of Psychology*, Vol. 41, pp. 417-440.
- Evans, C. & Cools, E. 2011. Applying styles research to educational practice. *Learning and Individual Differences*, 21, pp. 249-254
- Lescano G., Costaguta R. & Menini, M. 2016. Applying Data Mining to Discover Successful Collaborative Groups Styles. *In 8th Euro American Conference on Telematics and Information Systems (EATIS 2016)*, Cartagena de Indias, Colombia.
- Mesurado, B., Tortul, M. & Schönfeld, F. 2018. Cinco grandes rasgos de personalidad: su relación con el flujo y el compromiso académico, Big Five Traits Of Personality: Theirrelations With Flow And Academic Engagement.
- Morales Rojas, Y. & Rojas Pinto, J. 2016. La personalidad y el estilo de aprendizaje: una relación mutua en el rendimiento académico en la clase de inglés. *Revista de Lenguas Modernas*, Nro (24). DOI: <https://doi.org/10.15517/rlm.v0i24.24688>
- O'Connor, M. & Paunonen, S. 2007. Big Five Personality Predictors of Post-Secondary Academic Performance. *In Personality and Individual Differences*, Vol. 43 (5), pp, 971-990. DOI: <https://doi.org/10.1016/j.paid.2007.03.017>.
- Paz Penagos, H. 2014. Aprendizaje autónomo y estilo cognitivo: diseño didáctico, metodología y evaluación. *Revista Educación en Ingeniería*, 9(17), pp. 53-65.
- Soller, A. 2001. Supporting Social Interaction in an Intelligent Collaborative Learning System. *In International Journal of Artificial Intelligence in Educ.*, Vol. 12(1), pp. 40-62.
- Stahl, G., Koschmann, T., & Suthers, D. 2006. Computer-Supported Collaborative Learning. *En Sawyer, R. K. (ed.) The Cambridge Handbook of The Learning Sciences*, pp. 409-426. Cambridge, UK: Cambridge University Press.

Salazar, N. & Durán, E. 2016 Objetos de aprendizaje personalizados para entornos ubicuos. Una revisión de antecedentes. *En actas de XI Jornadas de Ciencia y Tecnologías de facultades de Ingeniería del NOA*. UNJU.

Torres-Acosta, N., Rodríguez Gómez, J., & Acosta-Vargas, M. 2013. Personalidad, aprendizaje y rendimiento académico en medicina, *En Investigación en Educación*

Médica, Vol 2, Issue 8, 2013, pp 193-201, DOI: [https://doi.org/10.1016/S2007-5057\(13\)72712-1](https://doi.org/10.1016/S2007-5057(13)72712-1).

Wu, W., Chen, L., Yang, Qi & Li, Y. 2019. Inferring Students' Personality from Their Communication Behavior. In *Web-based Learning Systems. Intern. Journal of Artificial Intelligence in Education*. DOI: <https://doi.org/29.10.1007/s40593-018-00173-9>.

Nuevos criterios para formación de grupos colaborativos en COLLAB

Camila Arce, Pablo Santana-Mansilla y Rosanna Costaguta

Instituto de Investigación en Informática y Sistemas de Información (IIISI)

Facultad de Ciencias Exactas y Tecnologías (FCEyT)

Universidad Nacional de Santiago del Estero (UNSE)

camilarce2710@gmail.com; psantana@unse.edu.ar; rosanna@unse.edu.ar

RESUMEN

En los sistemas de Aprendizaje Colaborativo Soportado por Computadora los estudiantes trabajan en grupos interactuando mediante el uso de computadoras. Para promover el éxito de las actividades que esos grupos desarrollan, es necesario que la conformación de esos grupos no se decida de manera aleatoria sino bajo determinados criterios.

Existen algunas investigaciones que demuestran que considerar el estilo de aprendizaje de los estudiantes puede ser la característica utilizada para agruparlos con determinados pares. Sin embargo, existen otras características que podrían considerarse para ello, como podrían ser los estilos de personalidad, entre otras.

Docentes y estudiantes de la Universidad Nacional de Santiago del Estero disponen desde hace varios años de la aplicación COLLAB. Esta aplicación fue especialmente creada para dar soporte a las actividades de diálogo en pequeños grupos de estudiantes colaborativos.

Esta línea de trabajo propone definir nuevos criterios para agrupar a los estudiantes a fin de formar grupos en sistemas de Aprendizaje Colaborativo Soportado por Computadora. La validación de los criterios que se definan será efectuada mediante su implementación en COLLAB como nuevas funcionalidades de la aplicación.

Palabras clave: *Aprendizaje Colaborativo Soportado por Computadora, formación de grupos, estudiantes colaborativos.*

CONTEXTO

En este artículo se presenta una de las líneas de investigación del proyecto “*Desarrollo de aplicaciones para colaboración en e-learning*”, acreditado y financiado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Santiago del Estero (SECyT - UNSE) para el período 2022-2025. El proyecto responde a una de las líneas prioritarias de investigación del IIISI, vinculada con la creación y uso de tecnologías informáticas destinadas a mejorar los procesos de enseñanza y de aprendizaje. Por otro lado, el proyecto continúa una línea de investigación iniciada en 2010-2011 por el proyecto 23/C089 “*Fundamentos Conceptuales y Soportes Tecnológicos de la Informática Educativa*”, y continuada en 2012-2016 por el proyecto 23/C098 “*Sistemas de Información Web Basados en Agentes para Promover el Aprendizaje Colaborativo Soportado por Computadora*”, y en 2017-2021 por el proyecto 23/C138 “*Mejorando escenarios de aprendizaje colaborativo soportado por computadora*”. Todos los proyectos citados fueron acreditados y financiados por SECyT – UNSE.

1. INTRODUCCIÓN

En las últimas décadas se potenció el uso de Internet como herramienta para la enseñanza y el aprendizaje, y como tal fue incluida en el área del Aprendizaje Colaborativo Soportado por Computadora (ACSC). En los últimos años, principalmente por efecto de la pandemia por Covid-19, se acentuó la necesidad de utilizar plataformas virtuales que favorezcan el ACSC.

Independientemente del necesario diseño de las actividades de aprendizaje colaborativo que se vayan a desarrollar en esas plataformas, para alcanzar un adecuado ACSC resulta importante agrupar a los estudiantes de manera tal que se propicien buenos desempeños individuales y grupales.

Actualmente, los docentes y estudiantes de la Universidad Nacional de Santiago del Estero disponen de una aplicación web especialmente creada para gestionar y soportar experiencias de aprendizaje colaborativo soportado por computadora basadas en diálogo (síncrono y asíncrono). Esta aplicación, denominada COLLAB [1], fue desarrollada años atrás por otra de las líneas de trabajo del proyecto de investigación al que pertenece la presente propuesta. COLLAB funciona en los servidores de la universidad para facilitar el acceso de la comunidad universitaria.

De acuerdo con el estudio realizado por Costaguta [2], los grupos en ACSC se conforman de diferentes maneras: en algunos casos es el profesor quien selecciona aleatoriamente a los miembros de cada grupo, o establece algún criterio para dicha selección; en otros, son los alumnos quienes se agrupan por amistad, afinidad en la forma de trabajar, etc.

En la aplicación COLLAB el docente dispone de funcionalidades que le permiten conformar los grupos de estudiantes seleccionándolos de manera directa, eligiéndolos de manera aleatoria, o creando

los grupos mediante un algoritmo genético que considera las mejores combinaciones de estudiantes en base a sus respectivos estilos de aprendizaje. Sin embargo, COLLAB no solo dispone del estilo de aprendizaje de cada estudiante, también cuenta con la identificación de su estilo de personalidad. A la fecha esta última característica no ha sido considerada para realizar agrupamientos en COLLAB, tampoco se ha probado con una combinación de ambos estilos.

Existen varios modelos de estilos de aprendizaje, pero el considerado en esta investigación es el ideado por Felder y Silverman [3], quienes proponen cuatro dimensiones de análisis y dos estilos contrapuestos para cada una de ellas: dimensión 1, percepción (estilos intuitivo o sensitivo), dimensión 2, entrada (estilos visual o verbal), dimensión 3, procesamiento (estilos activo o reflexivo), y dimensión 4, comprensión (estilos global o secuencial). Considerando las combinaciones de valores posibles para las cuatro dimensiones se pueden definir 16 estilos distintos. Así, un estudiante puede manifestar un estilo intuitivo, verbal, activo y secuencial mientras otro responder a un estilo sensitivo, visual, activo y global, o a cualquier otra combinación posible.

En la aplicación COLLAB, para determinar su estilo de aprendizaje, los estudiantes responden un cuestionario compuesto por cuarenta y cuatro preguntas creado por Felder y Soloman [4]. Puede accederse al cuestionario desde un enlace que se encuentra disponible en la interfaz principal.

De igual manera disponen de un enlace que les permite responder un cuestionario para descubrir sus estilos de personalidad considerando el modelo Big Five [5]. Este modelo establece cinco categorías:

- Neuroticismo: tendencia a experimentar emociones negativas de forma crónica como depresión, ansiedad e ira.
- Extroversión: predisposición a tener emociones positivas y rasgos relacionados con actividad y energía, con tendencia a ser sociables, activos, locuaces y positivos.
- Apertura: relacionada con creatividad científica y artística, pensamiento divergente, liberalismo político y búsqueda proactiva y apreciación de nuevas experiencias.
- Amabilidad: orientación prosocial con tendencia a ser considerados, compasivos, cooperativos, confiados, amigables, perdonadores, generosos, y dispuestos a comprometer sus intereses con los de los demás.
- Responsabilidad: tendencia a la organización, autocontrol, perseverancia y motivación en conductas dirigidas a metas.

El objetivo de esta investigación es ampliar las funcionalidades de COLLAB para que los docentes que lo usan puedan valerse de nuevos criterios aplicables en la identificación de los estudiantes que conformarán cada uno de los grupos de aprendizaje.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO/OBJETIVOS

Con este proyecto se busca dar respuesta a los siguientes interrogantes:

- ¿Pueden utilizarse los estilos de personalidad para agrupar a los estudiantes promoviendo la mejora de sus desempeños?
- ¿Pueden utilizarse de manera simultánea los estilos de aprendizaje y de personalidad para agrupar a los estudiantes promoviendo la mejora de sus desempeños?

- ¿Es mejor formar grupos homogéneos o heterogéneos?
- ¿Es posible automatizar en COLLAB la formación de grupos aplicando los nuevos criterios elaborados?

Para alcanzar estos objetivos, primero se realizará la búsqueda de bibliografía vinculada con aprendizaje colaborativo soportado por computadora, y criterios aplicados hasta el momento para la conformación de grupos colaborativos de estudiantes en contextos educativos. Luego, se analizarán los antecedentes recabados y se elaborarán nuevos criterios para la formación automática de grupos. Esos nuevos criterios se implementarán como nuevas funcionalidades en la aplicación COLLAB. Para finalizar, se validarán sus correctos funcionamientos mediante su uso por parte de estudiantes y docentes reales.

3. RESULTADOS ESPERADOS

Algunos de los resultados a los que se espera arribar al concretar este trabajo son:

- Estado del arte referido a los criterios implementados para conformación de grupos de estudiantes colaborativos en contextos educativos.
- Una ampliación de las actuales funcionalidades de la aplicación COLLAB, a fin de permitir identificar a los estudiantes que integrarán los grupos colaborativos considerando no sólo el estilo de aprendizaje.
- Resultados experimentales que permitan validar los criterios de agrupamiento definidos.
- Difusión de los resultados alcanzados mediante presentaciones en congresos y publicación de artículos en revistas.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo de esta línea de investigación está compuesto por un docente investigador formado, y un docente

investigador en formación. El equipo también cuenta con un integrante estudiante que está desarrollando su tesis de grado para obtener el título de Licenciado en Sistemas de Información en el marco de esta investigación.

5. REFERENCIAS

- [1] Lescano, G. & Costaguta, R. (2018). COLLAB: Conflicts and Sentiments in chats. In *Proceedings of the XIX International Conference on Human Computer Interaction (Interacción 2018)*. Association for Computing Machinery (ACM), New York, USA, Article 33, pp. 1–4. DOI: <https://doi.org/10.1145/3233824.3233864>
- [2] Costaguta, R. (2015). Algorithms and Machine Learning Techniques in

Collaborative Group Formation. *LNAI 9414: Advances in Artificial Intelligence and Its Applications*. Pichardo Lagunas, Obdulia, Herrera Alcántara, Oscar, Arroyo Figueroa, Gustavo (Eds.). pp. 249-258. DOI: https://doi.org/10.1007/978-3-319-27101-9_18

[3] Felder, R., & Silverman, L. (1988). Learning and Teaching Styles in Engineering Education Application. *Journal of Engineering Education*. Vol. 78(7).

[4] Felder, R., & Soloman, V. (1984). Index of Learning Styles. Disponible en: <https://www.webtools.ncsu.edu/learningstyles> Último acceso: 04/03/2022.

[5] Digman, J. (1990). Personality Structure: Emergence of the Five-Factor Model. *Annual Review of Psychology*, Vol. 41, pp. 417–440.

SISTEMA INFORMÁTICO DE PREDICCIÓN DE DESERCIÓN ESTUDIANTIL UNIVERSITARIA BASADO EN UN MODELO DE INDICADORES DE REGULACIÓN DEL APRENDIZAJE, EN ENTORNOS EDUCATIVOS MEDIADOS POR TIC

**ISTVAN, Romina; LASAGNA, Valeria;
BACIGALUPE, María de los Ángeles; RIVERO, Julieta**

*Universidad Tecnológica Nacional, Facultad Regional La Plata
Grupo de I&D Aplicado a Sistemas Informáticos, GIDAS UTN FRLP*

*Av. 60 s/n° esquina 124, CP 1900, La Plata, Buenos Aires, Argentina.
{ristvan; valerial; mabacigalupe}@frlp.utn.edu.ar; julietaa.rivero@gmail.com*

RESUMEN

La regulación del aprendizaje permite a los estudiantes transformar sus aptitudes mentales en competencias académicas convirtiéndose en un importante factor predictor del rendimiento académico y por consiguiente del logro estudiantil.

Incorporar este constructo a herramientas informáticas de gestión tutorial posibilita a las instituciones educativas caracterizar el perfil de cada estudiante y desarrollar estrategias de retención particularizadas promoviendo la tasa de graduación efectiva.

En este marco, el presente proyecto plantea como objetivo general incorporar al actual sistema de gestión tutorial de la UTN FRLP (Sistema ESDEU) indicadores de rendimiento académico, presencialidad y regulación del aprendizaje obtenidos de entornos virtuales de enseñanza-aprendizaje (EVEAS).

Palabras clave: Sistemas de Gestión Tutorial, Regulación del Aprendizaje, Deserción

Estudiantil, Rendimiento Académico, Tutorías, Indicadores de Deserción.

CONTEXTO

Particularmente, la UTN La Plata en el año 2017 comenzó a trabajar formalmente en el diseño y desarrollo del Sistema ESDEU dentro del marco del Proyecto de Investigación y Desarrollo (PID): «Estudio Sistemático de Deserción Estudiantil Universitaria» homologado por Rectorado UTN con código TEINNLP0003786, el cual continuó su línea de trabajo a partir del año 2020 con un nuevo PID: «Determinación de perfiles de riesgo de deserción estudiantil en UTN-FRLP utilizando técnicas de minería de datos» con código TEUTNLP0007653.

La función principal del Sistema ESDEU se basa en calcular automáticamente los factores de riesgo de deserción, presentar alertas tempranas y gestionar de manera eficiente el seguimiento de los estudiantes facilitando el conjunto de tareas de gestión tutorial.

Para realizar el cálculo del riesgo de deserción utiliza una fórmula matemática fundamentada en

los principales enfoques que conforman el marco teórico y el análisis particular del perfil de los estudiantes desertores de la institución conformando tres pilares o ejes de análisis:

(I) Perfil Socioeconómico.

(II) Inasistencias (Cursadas Libres y Faltas para el ciclo lectivo en curso).

(III) Rendimiento Académico (Promedio, Finales y Parciales).

Cada uno de estos ejes se asocian con una medida de importancia relativa o peso, que representa la incidencia sobre la estimación. A mayor peso, mayor incidencia.

La Estimación de Riesgo de Deserción obtenida para cada alumno en particular se visualiza en el sistema de gestión tutorial mediante un semáforo de colores: verde (sin riesgo), amarillo (riesgo leve), naranja (riesgo moderado) y rojo (crítico), junto a su valor numérico representativo.

El nuevo proyecto: “Sistema informático de predicción de deserción estudiantil universitaria basado en un modelo de indicadores de regulación del aprendizaje, en entornos educativos mediados por TIC” complementa el trabajo precedente y constituye junto con ellos en el primer estudio formal sobre deserción en la UTN La Plata, respondiendo a la necesidad de elevar la tasa de graduación efectiva en cada una de las especialidades que brinda la Regional.

1. INTRODUCCIÓN

La sociedad actual demanda nuevos perfiles profesionales capaces de intervenir en contextos laborales cambiantes, multidisciplinares, colaborativos, regionalizados y fuertemente mediados por las Tecnologías de la Información y la Comunicación (TIC). En este sentido, el sistema universitario argentino trabaja en la

necesidad de consolidar un modelo de aprendizaje centrado en el estudiante que incorpore los procesos de enseñanza mediados por la tecnología e incluya un enfoque basado en competencias.

En este contexto, el concepto de regulación del aprendizaje se transforma en un constructo esencial para responder a esta nueva demanda, ya que desarrolla la competencia de aprender a aprender, permitiendo a los estudiantes transformar sus aptitudes mentales en competencias académicas. Se convierte además en un importante factor predictor del logro estudiantil permitiendo a las instituciones educativas desarrollar políticas y estrategias curriculares y pedagógicas tendientes a minimizar la deserción estudiantil.

Motivado por las nuevas modalidades educativas por la pandemia relacionada con el covid-19 y teniendo en cuenta que el sistema ESDEU comenzó su desarrollo en el año 2017 donde las clases presenciales eran parte de la única modalidad brindada por la institución, se plantea con este nuevo proyecto la incorporación de indicadores del nuevo concepto de presencialidad, regulación del aprendizaje y rendimiento académico en entornos virtuales de enseñanza-aprendizaje (EVEAS).

El proyecto establece como objetivos específicos:

- Caracterizar el concepto de nueva presencialidad de los estudiantes de la UTN FRLP en los entornos virtuales de enseñanza-aprendizaje (EVEA).
- Caracterizar la regulación del aprendizaje de los estudiantes en entornos virtuales de enseñanza-aprendizaje y su relación con el

rendimiento académico.

- Determinar métricas que ayuden a medir la regulación del aprendizaje en ambientes mediados por TIC.
- Construir un modelo de indicadores que posibilite la medición de la regulación del aprendizaje en ambientes mediados por TIC.
- Incluir los procesos de metacognición en las métricas y modelo de indicadores previamente obtenidos para evaluar en forma holística los mecanismos de regulación del aprendizaje en dichos entornos.
- Determinar la relación entre regulación del aprendizaje, rendimiento académico y los principales factores causales de la deserción estudiantil universitaria en la institución.
- Incorporar los modelos obtenidos al actual sistema de gestión tutorial ESDEU.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

El área temática que el proyecto aborda responde fundamentalmente a investigación y desarrollo en *Tecnología Informática Aplicada en Educación*.

Particularmente trabaja con las líneas de investigación de *Educación a Distancia* y *Psicología Cognitiva* aplicada a Informática Educativa. Dentro de ellas aborda los constructos, métricas y sus relaciones en los entornos virtuales de enseñanza-aprendizaje (EVEA) de:

- *Nueva presencialidad*
- *Regulación del aprendizaje*
- *Rendimiento académico*
- *Deserción estudiantil*

3. RESULTADOS OBTENIDOS / ESPERADOS

Se espera como resultado final del proyecto incorporar al actual sistema de gestión tutorial ESDEU indicadores del desempeño académico obtenidos de entornos virtuales de enseñanza-aprendizaje (EVEAS).

Se espera con estas nuevas funcionalidades que la herramienta opere tanto en la modalidad presencial, virtual o mixta.

Atendiendo al objetivo general el proyecto contribuye a:

- 1- Promover las intervenciones de docentes y tutores facilitando la aplicación de estrategias tendientes a mejorar la regulación del aprendizaje y el desempeño académico de los alumnos en concordancia con el nuevo modelo propuesto por el Programa INNOVA@UTN de UTN Rectorado (2020). El modelo buscado se asocia con la enseñanza de la tecnología vinculada con la profesión desde la perspectiva teórica de una enseñanza centrada en el estudiante y el desarrollo de competencias propias del desempeño profesional.
- 2- Brindar un modelo de detección temprana de deserción estudiantil que contribuya a la mejora de los indicadores académicos en general.
- 3- Promover el incremento en las tasas de promoción efectiva, lo cual redundará directamente en la reducción de la deserción universitaria. De esta manera, contribuye con las Políticas Públicas de transformación social en el área de la Educación Superior.
- 4- Su aplicación particular en la UTN tiene como fin formar profesionales con una orientación científico-tecnológica, promoviendo de esta manera, una participación activa en el desarrollo de la Industria.
- 5- Visibilizar indicadores que sirvan de

comparación con otras instituciones ofreciendo un valioso aporte a la comunidad académica universitaria en general.

Se espera la aplicación y uso final a nivel institución.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está conformado por un Director, un Codirector, dos Docentes Investigadores de la Carrera de Ingeniería en Sistemas de Información (DISI) y tres alumnos becarios de investigación.

Cuenta con una tesis de Maestría y una de Especialización en la carrera de Tecnología Informática Aplicada en Educación, y tres desarrollos de Prácticas Supervisadas (PS) de los estudiantes, necesarias para la obtención del título de grado de la carrera de Ingeniería en Sistemas de Información.

5. BIBLIOGRAFÍA

- [1] Bernardo, A., Esteban, M., Cervero, A., Cerezo, R., & Herrero, F. J. (2019). The influence of self-regulation behaviors on university students' intentions of persistence. *Frontiers in psychology*, 2284.
- [2] Isohätälä, J., Järvenoja, H., & Järvelä, S. (2017). Socially shared regulation of learning and participation in social interaction in collaborative learning. *International Journal of Educational Research*, 81, 11-24.
- [3] Kahu, E. R., & Nelson, K. (2018). Student engagement in the educational interface: Understanding the mechanisms of student success. *Higher education research & development*, 37(1), 58-71.
- [4] Kuz, A., Falco, M., Nahuel, L. y Giandini,

R.S. "Agent SocialMetric: Una Aplicación Práctica de Solución TIC como Soporte a la Enseñanza". En Quinta Conferencia de Directores de Tecnología de Información, TICAL 2015. Gestión de las TICs para la Investigación y la Colaboración, 6 al 8 de julio de 2015. Viña del Mar, Chile.

[5] Morgan, T. (2020). Exploring socially shared regulated learning in PBL using group digital learning diaries: A study design. In DS 104: Proceedings of the 22nd International Conference on Engineering and Product Design Education (E&PDE 2020), VIA Design, VIA University in Herning, Denmark. 10th-11th September 2020.

[6] Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in psychology*, 8, 422.

[7] Programa Innova@UTN, Secretaría Académica, UTN Rectorado, 2020

[8] Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. A. (2020). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior*, 107, 105512.

[9] Tulabut, O. A. P., Mangalus, R. S., Cruz, R. N. C., Gonzales, A. B., Pare, E. L. L., Evaristo, I. M. C., & De Ala, C. L. M. (2021). Improving Academic Success Through Applied Learning Theories, A General Elective (GE) Course. *Higher education*, 18, 26.

[10] Wandler, J. B., & Imbriale, W. J. (2017). Promoting undergraduate student self-regulation in online learning environments. *Online Learning*, 21(2), n2.

[11] Winne, P. H. (2017). Learning analytics for self-regulated learning. Handbook of learning analytics, 241-249.

DISEÑO DE SISTEMAS RECOMENDADORES EN EL ÁMBITO DE LAS BIBLIOTECAS UNIVERSITARIAS

Sergio A. Cabrera, Cristian D. Pacifico y Juan C.L. Teze

Área de Agentes y Sistemas Inteligentes, Facultad de Ciencias de la Administración,
Universidad Nacional de Entre Ríos,
Monseñor Tavella 1424, (E3202KAC) Concordia

{sergio.cabrera, cristian.pacifico, carlos.teze}@uner.edu.ar

RESÚMEN

Esta línea de I+D pretende diseñar un modelo de sistema que utilice técnicas de recomendación y de Machine Learning, para realizar sugerencias de fuentes de información en el ámbito de las bibliotecas universitarias. Estas técnicas podrían integrarse con el trabajo diario del bibliotecario, permitiendo potenciar los servicios y las funciones de las bibliotecas.

Palabras clave: Sistemas Recomendadores, Bibliotecas Universitarias, Machine Learning.

CONTEXTO

La línea de I+D se ubica dentro del grupo de investigación *Agentes y Sistemas Inteligentes* de la Facultad de Ciencias de la Administración dependiente de la Universidad Nacional de Entre Ríos.

Persigue como propósito diseñar un modelo de sistema que utilice técnicas de recomendación y de aprendizaje automático, para realizar sugerencias de fuentes de información, en el ámbito de las bibliotecas universitarias.

INTRODUCCIÓN

La biblioteca es una organización que tiene como misión fundamental la de otorgar acceso a la información a diferentes usuarios que hacen uso de su acervo documental.

Las bibliotecas universitarias se caracterizan por brindar acceso y difusión de fuentes de información, colaborando en los procesos de

creación de conocimiento, sirviendo de apoyo a la educación y a la investigación, tanto a la comunidad académica en la que se encuentra inmersa, como así también a la sociedad en general.

Se entiende por fuente de información a todo objeto o sujeto que genere, contenga, suministre o transfiera información.

Los cambios que se han ido generando en las bibliotecas, tanto como las tecnologías incorporadas, permiten ofrecer a los usuarios, información actualizada en el momento que la necesitan.

Ponjuán Dante [1], señala que las transformaciones tecnológicas ocurren a velocidades insospechadas, lo que impone, muchas veces, un tiempo de transición en el cual coexisten los nuevos y los antiguos sistemas.

La aplicación de la tecnología computacional a las etapas de la actividad científico informativa, ha favorecido la utilización de los Sistemas Integrados de Gestión Bibliotecaria (SIGB).

Se puede definir un SIGB, como un conjunto de programas informáticos diseñados para atender los procesos de las bibliotecas, brindando soluciones a los requerimientos de las mismas, ayudando en los procesos técnicos y en los servicios prestados, favoreciendo la intercomunicación entre bibliotecas con el objetivo de conformar catálogos colectivos.

Dentro de estos sistemas se destaca KOHA, por ser el primer SIGB de software libre. Fue inicialmente desarrollado en 1999, y la primera biblioteca con este sistema surge en enero del 2000. Incluye la administración de

los registros internacionales y las normas de catalogación, utilizando el formato MARC y el protocolo Z39.50 para la recuperación de información. Fue creado en Nueva Zelanda por Horowhenua Library Trust y Katipo Communications Ltd. [2].

Existen varias iniciativas que proponen la utilización de la Inteligencia Artificial (IA) y en especial de técnicas Aprendizaje Automático (ML), en plataformas y aplicaciones destinadas a bibliotecas.

Según Russel [3], en el aprendizaje de máquinas una computadora observa datos, construye un modelo basado en esos datos y utiliza ese modelo simultáneamente, como una hipótesis acerca del mundo y como un elemento de software que puede resolver problemas.

En este contexto, es necesario la utilización de mecanismos que permitan extraer información relevante de los conjuntos de datos, porque los datos por sí mismos no proporcionan mucha información. Estos se deben procesar, organizar, estructurar o presentar en un contexto dado para hacerlos realmente útiles.

Según Maes, Jannach, Ricci y Melville, como se citó en Briguez [4], los sistemas de recomendación son mecanismos de apoyo que ayudan a los usuarios en su proceso de toma de decisiones al interactuar con volúmenes de información grandes o complejos. La mayoría de los sistemas de recomendación están destinados a ayudar a los usuarios para hacer frente al problema de la sobrecarga de información, facilitando el acceso a los elementos pertinentes.

Los recomendadores intentan generar un modelo del usuario o de la tarea del usuario y aplicar diversas heurísticas para anticipar qué tipo de información puede ser útil. De esta manera, tienen la capacidad de predecir si un elemento en particular puede ser de interés para un usuario.

Frente a la gran cantidad y variedad de fuentes de información existentes en el ámbito educativo, realizar la selección del material apropiado en una biblioteca

universitaria, es una tarea cada vez más difícil.

La complejidad en la selección se manifiesta en la tarea diaria del bibliotecario, como así también en las necesidades de los usuarios de la biblioteca. En el primer caso, el bibliotecario debe asegurar que la colección refleje los intereses de sus usuarios, mientras que los mismos deben satisfacer en ella sus demandas informacionales generadas a través de sus consultas.

Ricci, como se citó en Charnelli [5], establece que la tarea de un Sistema Recomendador (SR) es seleccionar automáticamente los artículos más apropiados para cada usuario de acuerdo a sus intereses y preferencias personales. Generalmente, un SR se centra en un tipo específico de elemento a recomendar denominado "ítem", como por ejemplo un repositorio de materiales educativos, que genera recomendaciones personalizadas para proporcionar sugerencias útiles y efectivas para ese tipo específico de ítem.

Se debe tener en cuenta que dependiendo del enfoque de recomendación, según el contexto y la necesidad del usuario, los SR generan sugerencias utilizando diversos tipos de conocimiento y datos sobre los usuarios, los ítems disponibles y las transacciones anteriores almacenadas en bases de datos personalizadas.

En este sentido, la Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA) [6], ha enunciado una serie de políticas y recomendaciones con el objetivo de redefinir el trabajo de los bibliotecarios y el rol que las bibliotecas desempeñan en una sociedad que cada día más se amalgama con la IA y en especial con las técnicas de ML. Las técnicas de IA y ML podrían integrarse con el trabajo diario del bibliotecario, permitiendo potenciar los servicios y las funciones de las bibliotecas.

LÍNEAS DE I+D

La línea de I+D se vincula con *Tecnología Informática Aplicada en Educación*, pretende

desarrollar herramientas para la selección de fuentes de información en bibliotecas universitarias.

Si bien los SR son objeto de estudio de varias investigaciones y tienen aplicación en numerosos dominios, como por ejemplo el comercio electrónico, recién a principios del año 2000 comenzaron a aparecer las primeras aplicaciones destacadas en el campo de la educación.

Los sistemas recomendadores aplicados a la educación (SRE) pueden ser muy diversos. La mayor parte de ellos se utiliza para sugerir recursos de aprendizaje o para sugerir secuencias de recursos. En el área de investigación, los SRE se han estudiado principalmente como un medio para asistir automáticamente a usuarios en la búsqueda de recursos adecuados [5].

Es posible que el usuario pueda examinar las recomendaciones, las pueda aceptar y además suministrar retroalimentación de manera implícita o explícita. Es probable almacenar todas estas acciones del usuario en la base de datos del recomendador, con el objetivo de utilizarlas para generar nuevas sugerencias en futuras interacciones del usuario con el sistema.

Estas técnicas de recomendación pueden mejorarse con las desarrolladas para ML. Los avances en este sentido agregarían nuevas dimensiones y enfoques a los procesos de gestión del conocimiento que se producen en las bibliotecas, en particular, los referidos a la organización, el almacenamiento y la integración de conocimiento.

Entre algunos desarrollos actuales, es posible citar a Ex Libris Alma [7], como una plataforma unificada de servicios de biblioteca basada en la nube que brinda soporte para la adquisición, la administración y el préstamo de recursos impresos, electrónicos y digitales. DARA es el motor inteligente que da soporte a las decisiones de Alma, dotando de IA a las bibliotecas; utiliza algoritmos y análisis inteligentes para adaptar sus recomendaciones a las características de cada biblioteca.

Además cabe mencionar a Hamlet, que es un sistema que utiliza ML promoviendo interfaces exploratorias experimentales para la colección de Tesis del Instituto Tecnológico de Massachusetts (MIT). Está en desarrollo y algunas interfaces no se han finalizado todavía. [8][9].

Por el momento, no se conocen publicaciones de estudios comparativos de estos frameworks y herramientas; como así tampoco del grado de integración con los SIGB tradicionales.

En este proceso de investigación, está previsto:

- Especificar a los SIGB.
- Caracterizar las técnicas de recomendación.
- Particularizar la función de recomendación.
- Caracterizar las técnicas de aprendizaje automático para recomendación.
- Diseñar un modelo de arquitectura para sistemas recomendadores
- Realizar un análisis comparativo de las soluciones existentes.

RESULTADOS ESPERADOS

De la especificación de los SIGB, se identificarán las funcionalidades y los componentes esenciales, resultando de especial interés el estudio de las funciones para el servicio de referencia y el de diseminación de la información.

De la caracterización de las técnicas de recomendación, se espera lograr una síntesis de las técnicas de recomendación que asocie a cada una con sus dominios de aplicación más adecuados.

Al particularizar la función de recomendación, se espera obtener un esquema de requisitos y de características deseables de un sistema para la recomendación de fuentes de información. Luego se realizará una síntesis de las técnicas de aprendizaje automático y su forma de aplicación en el contexto de recomendación de fuentes de información.

A estas instancias, se estará en condiciones de diseñar y especificar una arquitectura general para sistemas recomendadores, proponiendo una guía metodológica para su implementación.

Para finalizar se pretende obtener una matriz comparativa de ventajas y desventajas de las soluciones propuestas.

FORMACIÓN DE RRHH

La línea de I+D se llevará adelante dentro de un equipo de investigación en el área *Agentes y Sistemas Inteligentes* de la FCAD/UNER. Esto permitirá que la experiencia de un grupo de investigación consolidado en el área, sirva para la formación de recursos humanos en la FCAD/UNER, debido a que en paralelo al desarrollo de la investigación, se llevará a cabo la redacción de una tesis de Maestría y una Tesina de grado

BIBLIOGRAFÍA

- [1] Ponjuán Dante, G. (2002). Biblioteca digital... Un nuevo paso en la evolución de las arquitecturas de información. *Ciencias de la Información*. 33(1), 55-63.
- [2] KOHA. Consultado el 10 de febrero de 2022. <https://koha-community.org/>
- [3] Russell, S. y Norvig, P. (2021). *Artificial Intelligence: A Modern Approach*. Pearson.
- [4] Briguez, C. E. *Integración de técnicas cualitativas y cuantitativas en los sistemas de recomendación*. [Tesis de Doctor en Ciencias de la Computación, Universidad Nacional del Sur]. Repositorio institucional <https://repositoriodigital.uns.edu.ar/bitstream/handle/123456789/4736/BRIGUEZ%20Cristian%20Emanuel%20-%20Tesis%20Doctoral.pdf?sequence=5&isAllowed=y>
- [5] Charnelli, M. E. (2019). *Sistemas recomendadores aplicados en educación*. [Tesis Especialización en Tecnología Informática Aplicada en Educación, Universidad Nacional de La Plata]. Repositorio institucional <http://sedici.unlp.edu.ar/handle/10915/67063>
- [6] IFLA (2020). *IFLA Statement on Libraries and Artificial Intelligence*. <https://repository.ifla.org/handle/123456789/1646>
- [7] ExLibris. Consultado el 10 de febrero de 2022. <https://exlibrisgroup.com/>
- [8] American Library Association. (2019). Artificial Intelligence and Machine Learning in Libraries. *Library Technology Reports*, 55(1). <https://doi.org/10.5860/ltr.55n1>
- [9] Yelton, A. (s.f.). *Hamlet*. Consultado el 18 de febrero de 2022. <https://hamlet.andromedayelton.com/>

Aproximación al diseño de aprendizaje aplicado a un curso de e-learning

Iris Sattolo, Marisa Panizzi

Universidad de Morón

Cabildo 134, Buenos Aires, Argentina

isattolo@unimoron.edu.ar, marisapanizzi@outlook.com

Resumen

La minería de datos educativos y la analítica del aprendizaje (LA) prometen una mejor comprensión del comportamiento y del conocimiento de los estudiantes. Los datos recopilados en los cursos de e-learning y las herramientas utilizadas por la analítica del aprendizaje no pueden, por sí solos, ofrecer beneficios para mejorar el diseño de aprendizaje (LD). Este documento presenta la línea de investigación y el trabajo en curso, que se está llevando a cabo en nuestra institución, sobre la sinergia que debe existir entre estos dos conceptos. El objetivo de la investigación actual es analizar los modelos y metodologías preexistentes para proporcionar un método que se ajuste al diseño de aprendizaje en un curso virtual.

Palabras clave: Minería de datos Educativa, Analítica del Aprendizaje, Diseños de Aprendizaje, e-learning.

Contexto

La línea de investigación que se reporta en este artículo es financiada por el proyecto de investigación titulado “Aplicación de Analítica del Aprendizaje sobre un curso a distancia desarrollado con técnicas de Diseño del Aprendizaje” (Código 80020190300011 UM), que fuera aprobado por la Secretaría de Ciencia y Tecnología de la Universidad de Morón.

Introducción

Es indiscutible que, transcurridos 22 años del

siglo XXI, la educación a distancia se ha instalado en la sociedad. La pandemia generada por el SAR-COV2 aceleró la implementación de la modalidad, irrumpiendo en todas las áreas de la comunidad educativa (clases, administrativos, directivos, etc.). El avance de la tecnología e internet posibilitó la comunicación entre los actores, logrando que las actividades no cesaran.

La Universidad de Morón cuenta con un sistema de gestión del conocimiento (en inglés, *Learning Management System*, LMS), también llamados (*Virtual Learning Environment* o VLE) o Entornos Virtuales de Aprendizaje (EVA) [1]. La plataforma utilizada en estos últimos 2 años fue Blackboard [2]. Estos sistemas de gestión del conocimiento generan datos que pueden utilizarse para mejorar la interacción del entorno con sus actores. La recopilación de estos datos se enmarca en la subdisciplina de la informática: la minería de datos (en inglés *Data Mining* o DM) que permite descubrir información nueva y potencialmente útil en grandes cantidades de datos [3].

La DM aplicada al ámbito educativo dio lugar a la aparición de la minería de datos educativos (en inglés, *Educational Data Mining* o EDM) la cual propone métodos para extraer información útil a partir de los datos generados en los entornos educativos [4]. Emparentada a la EDM, [5] en el año 2011 se comenzó a aplicar la analítica del aprendizaje en los centros universitarios. El concepto de analítica del aprendizaje, LA por sus siglas en inglés - *Learning Analytic*- representa un campo multidisciplinario que involucra diversas áreas: *maching learning*, inteligencia artificial,

recuperación de información, estadísticas y visualización. Además, contiene las áreas de investigación como EDM, sistemas de recomendación y aprendizaje adaptable personalizado [6]. Este documento emplea la definición de análisis de aprendizaje establecida en la primera conferencia internacional sobre análisis de aprendizaje y conocimiento LAK 11 y adoptada por la Sociedad para la Investigación de análisis de aprendizaje (SoLAR) como: “la analítica de aprendizaje es la medición, recopilación, análisis e informe de datos sobre los alumnos y sus contextos, con el fin de comprender y optimizar el aprendizaje y los entornos en los que ocurre”.

En dicha conferencia también se comentó de la relación que debería tener el LA y el diseño de aprendizaje [7]. Estos autores entienden como diseños de aprendizaje (o como pueden denominarse modelos o patrones pedagógicos) a representaciones de la práctica docente. Los diseños de aprendizaje articulan una secuencia de actividades de aprendizaje, los recursos que utilizan los alumnos al participar en esas actividades y cómo el maestro u otros participantes en la experiencia apoyan las actividades de aprendizaje. Un diseño de aprendizaje puede actuar como un modelo o plantilla que los maestros pueden replicar en una variedad de contextos educativos. En esencia, los diseños de aprendizaje ayudan a enmarcar la intención y el proceso de la experiencia pedagógica [8]. El análisis de aprendizaje se afirma, permitiría probar los supuestos con datos reales de interacción del estudiante en lugar de medidas de autoinforme, como encuestas post hoc.

El diseño de aprendizaje (LD por sus siglas en inglés *Learning Designs*) es el proceso en el cual se realizan las siguientes actividades [9]:

- 1 definir las tareas,
- 2 proporcionar los contextos y recursos para realizar dichas tareas,
- 3 apoyar al alumno durante el desempeño de la tarea y

- 4 proporcionar retroalimentación sobre los resultados.

La investigación se ha centrado en “conceptualizar los principios del diseño del aprendizaje, sin evaluar lo que sucede después del proceso de diseño” [10].

En [10] se argumenta que el potencial evaluativo de la analítica de aprendizaje sería significativamente mejorado por referencia al diseño de aprendizaje que documenta la intención pedagógica, se identifica, como paso esencial para mejorar la efectividad de la evaluación y construir bases para los sistemas de recomendación, que se debe establecer un marco conceptual para los patrones de análisis de aprendizaje típicos esperados de diseños de aprendizaje particulares.

Investigadores, tales como en [10], enfatizan que la exploración futura debería considerar desarrollar un marco sobre cómo capturar y sistematizar datos de diseño de aprendizaje basados en el análisis de aprendizaje y teoría de aprendizaje, y documentar qué opciones de diseño de aprendizaje hechas por educadores influyen en actividades de aprendizaje y actuaciones a lo largo del tiempo. Los mismos han observado que, en la actualidad, el diseño del aprendizaje es muy diverso, porque la forma en que se conceptualiza el término depende de la elección de la perspectiva del observador. No obstante, el diseño del aprendizaje debe conceptualizarse antes de que pueda utilizarse como un proceso que conduzca a resultados de diseño explícitos y compartibles para el aprendizaje.

Nuestro grupo de investigación ha comenzado a trabajar sobre estos temas elaborando un estado del arte sobre la relación existente entre estos dos conceptos LA y LD [13]. Hemos evidenciado que existen distintas propuestas en torno al diseño de aprendizaje: se plantean modelos, *frameworks*, herramientas, como también se analizan distintas taxonomías desde distintas propuestas pedagógicas teóricas.

La construcción del estado de arte sobre minería de datos educacional, analítica del aprendizaje y diseño de aprendizaje aplicado al e-Learning (puntos 1 y 2 del plan de trabajo) se desarrolló mediante el método de investigación riguroso, mapeo sistemático de la literatura (en inglés, *Systematic Mapped Study*) [11].

Por todo lo expuesto anteriormente y continuando con nuestra propuesta donde se planteó la sinergia que debe existir entre la analítica del aprendizaje y el diseño del aprendizaje [14], presentamos las preguntas de investigación (PI) que guían nuestro trabajo en la siguiente sección.

Líneas de Investigación, Desarrollo e Innovación

Dentro del campo de la minería de datos educativos, analítica del aprendizaje, y diseños de aprendizaje, y en el contexto de los cursos en línea, seguimos orientando los esfuerzos en responder a estas preguntas de investigación:

PI 1: ¿Cuál fue el impacto originado al abordar el e-learning en el año 2020 en el dictado de las clases de los docentes de la Universidad de Morón?

PI 2: ¿Qué se debe tener en cuenta al realizar el diseño de un curso a dictarse a distancia, para que éste pueda ofrecer datos valiosos recuperables a través de los LMS?

PI 3: ¿Qué se debe tener en cuenta cuando utilizamos datos de los estudiantes para que no se viole la privacidad del usuario?

PI 4: Una vez realizado el análisis de los datos, luego de aplicar LA, los resultados obtenidos, ¿contribuyen a los resultados esperados?

Para dar respuesta a las preguntas de investigación planteadas hemos propuesto el siguiente plan de trabajo del cual ya se han resuelto los puntos 1, 2 y 3:

1) Elaboración del estado de arte de la Minería de Datos Educacional y Analítica del Aprendizaje.

- 2) Elaboración del estado del arte del Diseño de Aprendizaje aplicado al e-Learning.
- 3) Construir un instrumento de recolección de datos para responder la PI 1.
- 4) Analizar las respuestas aportadas por los docentes de la UM y recolectadas por el instrumento construido.
- 5) Analizar las metodologías existentes sugeridas por los trabajos seleccionados en el Estado del Arte en el abordaje del LD.
- 6) Proporcionar una metodología para diseñar un proyecto de curso virtual, evaluarlo y seguir su realización.
- 7) Aplicar técnicas de EDM y LA.
- 8) Puesta en marcha del curso y evaluación.

Para el diseño del instrumento de recolección de datos (punto 3 del plan de trabajo) se aplicaron las directrices del proceso de encuesta propuesto en [12] y [15].

Para el proceso de minería de datos (punto 7 del plan de trabajo) se aplicará el proceso “Descubrimiento de conocimiento en bases de datos” (en inglés *Knowledge Discovery in Databases o KDD*) [3].

Objetivos y Resultados Esperados

En esta línea de investigación en progreso se han logrado una serie de **resultados** que se detallan a continuación:

- a) Proyectos de Investigación: ha finalizado el PIO/17-01-MP-001 titulado: “Aplicación de tecnologías inteligentes de explotación de información para el análisis de perfiles de tesis de grado de carreras informáticas de la UM” (período 2017-2019). Proyecto en ejecución titulado: “Aplicación de Analítica del Aprendizaje sobre un curso a distancia desarrollado con técnicas de Diseño del Aprendizaje” (Código 80020190300011 UM) (período 2020-2022).
- b) Resultados Académicos, se han logrado dos tesis de grado correspondientes una a cada proyecto de investigación. Se desarrolla en el

marco del proyecto de investigación orientado, otra tesis de grado.

- c) Producción Científica: se ha presentado la línea de investigación en eventos científicos de alcance nacional (TE&ET¹ ediciones 2018 y 2019, CACIC² ediciones 2018 y 2019 y en el ámbito internacional, JENUI³ 2018 e InNGENIO⁴ 2019, 2020, 2021). Además, se logró una publicación en Springer.
- d) Formación en investigación: el grupo de investigación se encuentra en un proceso de aprendizaje constante de métodos de investigación de ingeniería de software experimental, revisiones sistemáticas [11] y encuestas [12], [15]. Además, se experimenta con procesos, técnicas y algoritmos de DM.

El grupo de investigación se plantea como **próximos objetivos**:

- Analizar los resultados del instrumento de recolección de datos con el propósito de identificar cómo impactó en los docentes de la UM el cambio en la modalidad del dictado de sus clases.
- Proporcionar y probar un modelo para ser aplicado en el diseño de aprendizaje.
- Presentar los avances y resultados experimentales en InNGENIO 2022, TEyET 2022 y CACIC 2022.

Formación de Recursos Humanos

El grupo se encuentra conformado por un Director, un Co-Director, un tesista de grado, y un docente-investigador en proceso de formación.

Se estima una tesis de la carrera Licenciatura en Sistemas en el marco de la línea de investigación.

Bibliografía

- [1] Belloc C. Entornos virtuales de Aprendizaje. Obtenido de: http://www.formaciondocente.com.mx/04_Rinc onTecnologia/03_AmbientesVirtuales/Entornos %20Virtuales%20de%20Aprendizaje%203.pdf
- [2] Blackboard: Educational Technology Services <https://www.blackboard.com/>
- [3] Hernández Orallo J., Ramírez Quintana M., Ferri Ramírez C. (2014). *Introducción a la Minería de Datos*. Ed. Pearson Educación S.A. Madrid.
- [4] Peña Ayala A. (2014). Educational Data Mining. A Survey and a data mining-based analysis of recent works. *Expert Systems with applications*, 41(4), pp. 1432-1446.
- [5] Siemens y Baker. Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. *LAK '12: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* April 2012 p.p. 252–254 doi.org/10.1145/2330601.2330661
- [6] Vahdat, M; Ghio, A; Oneto, L; Anguita, D; Funk, M; Rautherberg, M. (2015). Advances in Learning Analytics and Educational Data Mining. *ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. pp. 297-307.
- [7] Lockyer, L., Dawson, S. (2011, February). Learning designs and learning analytics. *In LAK'11 Proceedings of the 1st international conference on learning analytics and knowledge* (pp. 153-156).
- [8] Koper R., Bennett, S. (2008). Learning Design: Concepts. In H. Adelsberger, J. M.

¹ Congreso de “Tecnología en Educación” y “Educación en Tecnología” (TE&ET).

² Congreso Argentino de Ciencias de la Computación (CACIC).

³ Jornadas sobre la Enseñanza Universitaria de la Informática (JENUI 2018). Link: <http://jenui2018.uoc.edu/>

⁴ Congreso Latinoamericano de Ingeniería. Link: <http://fundacioniai.org/ingenio/>

Handbook on Information Technologies for Education and Training. Heidelberg: Springer Science & Business Media.

[9] Lockyer, L., Heathcote, E., Dawson, S. (2013). Informing pedagogical action: aligning learning analytics with learning design. *American Behavioral Scientist*, 57(10), pp 1439-1459.

[10] Mangaroska, K; Giannakos, M. (2019). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*. pp 516-534

[11] Kitchenham, B. y Chartes, S. (2007). *Guidelines for performing systematic literature reviews in Software engineering*, Keele University, EBSE-2007-01.

[12] Molléri J., Petersen K., Mendes E. (2020). An empirically evaluated checklist for surveys in software engineering. *Information and Software Technology* 119, 10624

[13] Sattolo I, Panizzi M. Contreras V, Valensise M. (2021). Análisis de la sinergia entre la Analítica del Aprendizaje y el Diseño del Aprendizaje. Estudio de Mapeo Sistemático. *Revista Desarrollo e innovación en Ingeniería* 2(6), pp 405- 414. ISBN 978-958-53278-6-3

[14] Panizzi M., Sattolo I. Analítica del aprendizaje y diseño del aprendizaje. (2021). *Workshop de investigadores en Ciencias de la Computación (WICC '2021)*, pp. 969-972.

[15] Kitchenham, B. A., & Pfleeger, S. L. (2008). Personal opinion surveys. In *Guide to advanced empirical software engineering* (pp. 63-92). Springer, London.

ESTRECHANDO EL CONTACTO ENTRE UNIVERSIDADES Y ESTUDIANTES: COMUNICACIÓN ANTE POSIBLES CASOS DE ABANDONO, PROPUESTAS PARA LA INSCRIPCIÓN. LÍNEA DE INVESTIGACIÓN Y DESARROLLO DEL CIDIA.

Martin Pustilnik^{1,2}, Monica Giuliano^{1,2}, Fernando Puricelli^{1,2}, Carlos Lombardi^{1,2}, Gerardo González Tulián^{1,2}, Federico Pagliari¹, Cristian Saldivia¹, Jonathan Ybarra¹, Magalí Gaiani¹

¹ *CIDIA: Centro de Investigación y Desarrollo en Informática Aplicada, Universidad Nacional de Hurlingham*

² *Instituto de Ingenierías y Tecnologías, Universidad Nacional de Hurlingham*

{martin.pustilnik; monica.giuliano; fernando.puricelli; carlos.lombardi}@unahur.edu.ar; gerardo.gonzaleztulian@gmail.com

RESUMEN

En la Universidad Nacional de Hurlingham (UNAHUR) se lleva a cabo desde 2019 un proyecto que busca disminuir el abandono en su población estudiantil.

El mismo tiene dos líneas de trabajo. La primera es generar un “Sistema de recomendación”, a través de un software que realiza recomendaciones a los estudiantes de materias a cursar, a partir de la historia académica individuales.

Una segunda línea de trabajo consiste en el estudio de las características de los estudiantes, para la identificación de indicadores en riesgo de abandono, mediante la aplicación de ciencia de datos. Se toma como información todas las bases de datos de la UNAHUR y se busca predecir el abandono de manera temprana para intervenir y asistir a los alumnos antes de que se produzca.

En 2021 se han alcanzado los primeros resultados, aunque preliminares, a partir de los datos disponibles en el SIU- Guarani. Se generaron satisfactoriamente mensajes personalizados para los estudiantes con

recomendaciones de cursada. Además, se elaboraron modelos de predicción de abandono con resultados promisorios.

Palabras clave: *Abandono, universidad, sistema de recomendación, ciencia de datos.*

CONTEXTO

El Centro de Investigación y Desarrollo en Informática Aplicada (CIDIA) se creó en 2019 con la intención de vincular investigadores y alumnos (becarios) de la Universidad Nacional de Hurlingham (UNAHUR).

Uno de los primeros proyectos impulsados por el CIDIA fue el proyecto de investigación “Estrechando el contacto entre universidades estudiantes: comunicación ante posibles casos de deserción, propuestas para la inscripción”, aprobado en la convocatoria PIUNAHUR 6¹. Dicho proyecto fue aceptado para su inclusión en el Banco Nacional de Proyectos de Desarrollo Tecnológico y Social (PDTs) del Ministerio de Ciencia, Tecnología e Innovación de la Nación, según Resolución 2021-1010-APN-SACT#MCT de la Secretaría

¹ <https://unahur.edu.ar/se-aprobaron-proyectos-de-la-convocatoria-piunahur6/>

de Articulación Científico-Tecnológica de dicho Ministerio.

El proyecto que aquí se presenta, se realiza con colaboración y asistencia de la Secretaría Académica y el área de Orientación Estudiantil. El proyecto resulta un punto de encuentro entre distintas áreas de la UNAHUR. En particular, esta iniciativa fomenta la integración de la comunidad de Informática dentro del ámbito de la UNAHUR.

1. INTRODUCCIÓN

En varias universidades nacionales, interesa contar con una gestión que asista y acompañe a cada estudiante en su trayectoria académica. En general, se busca garantizar en la práctica el derecho a la educación y propender al éxito de la mayor cantidad de estudiantes, atendiendo los desafíos que se derivan de las características de la población estudiantil. [1]

Entendemos que el abandono estudiantil es, tal vez, el factor individual que conspira en mayor medida contra el establecimiento de la educación universitaria como un derecho de nuestros jóvenes y adultos, y contra el rol de la universidad como un motorizador de movilidad social para los sectores menos favorecidos. Este fenómeno se manifiesta en forma generalizada en la mayor parte de las universidades públicas, y en forma particularmente aguda en las del conurbano bonaerense. [1]

Una de las características más relevantes es el crecimiento significativo de la población estudiantil, que en algunas universidades reviste características explosivas. Otra característica relevante es que una amplia proporción de los estudiantes cuenta con poco, o nulo, conocimiento sobre el funcionamiento de una universidad, y sobre lo que implica cursar una carrera de nivel universitario. Este fenómeno se deriva, al menos en parte, de la

gran cantidad de estudiantes que son primera generación de universitarios en su familia. [1-2]

Las dos características mencionadas están fuertemente presentes, en particular, en las universidades del conurbano bonaerense, casi todas de reciente creación. En particular, la UNAHUR fue creada el 2 de diciembre de 2014 [2-3].

En el caso puntual de la UNAHUR, el aumento explosivo de la matrícula se evidencia en casi 5.000 estudiantes en 2017 y más de 30.000 en 2021. Entre los múltiples desafíos que implica contar con una población estudiantil con estos rasgos, destacamos dos en los que focalizamos.

El primero es la gran cantidad de estudiantes que abandonan sus estudios, especialmente antes de completar el primer año de la carrera elegida. Entre las razones de este fenómeno, mencionamos la frustración que genera un bajo desempeño inicial, y la dificultad por adquirir los hábitos necesarios para transitar una carrera universitaria con altas chances de éxito. [1;5]

El segundo rasgo es la complejidad que reviste la definición de la oferta de cursos en la universidad. Uno de los motivos es que muchos estudiantes se inscriben en más materias, de las que su situación objetiva les permite cursar correctamente. Esto provoca altas cifras de abandono de cursos, y eventualmente también abandono de los estudios; por lo que hay una vinculación entre las dos cuestiones elegidas.

Estas dos problemáticas pueden ser atenuadas mediante la comunicación directa de distintos actores de la universidad (entre ellos docentes, directivos, personal ligado a la gestión académica, tutores) con cada estudiante, de modo de generar un vínculo de cercanía de los estudiantes con la universidad.

Por otro lado, el fenómeno de crecimiento mencionado al principio atenta contra la posibilidad de mantener una comunicación fluida, ya que a medida que la población estudiantil crece, no es acompañado por un crecimiento análogo del personal que puede participar en el fortalecimiento del vínculo.

Las TIC (Tecnologías de la Información y la Comunicación), se configuran en una oportunidad para el aprovechamiento de las potencialidades, en un sentido positivo desde el punto de vista socioeducativo.

Un recurso muy valioso al respecto es la gran cantidad de información acumulada sobre la trayectoria y el comportamiento de cada estudiante, en el sistema SIU-Guaraní y en la plataforma de acompañamiento al aprendizaje, Moodle. El análisis de esta masa de información, usando técnicas de ciencia de datos, puede asistir a la gestión académica en varios aspectos. Asistir en la detección de estudiantes que estén en riesgo de abandonar sus estudios o que los han abandonado recientemente, y también en la generación de propuestas de inscripción personalizadas para cada estudiante, basadas en el historial del rendimiento académico.

Adicionalmente, la posibilidad de enviar mensajes mediante redes sociales, el email o al celular, aumentaría en gran medida la capacidad de una institución universitaria para comunicarse efectivamente con su población estudiantil.

El proyecto que presentamos propone construir una solución informática que, explotando las potencialidades de las TIC, que brinde un mecanismo alternativo para mantener una comunicación fluida entre institución y estudiantes, que ayude en el abordaje de problemáticas específicas de la población estudiantil.

Un primer objetivo fue generar un software que brinde herramientas para potenciar la comunicación de la Universidad con sus estudiantes, generando mensajes que puedan llegar a una gran cantidad de estudiantes, con contenidos personalizados de acuerdo a la trayectoria académica de cada uno.

El segundo objetivo fue investigar el fenómeno del abandono en sí mismo en las distintas carreras de la UNAHUR.

El **objetivo** general consiste en definir y desarrollar aplicaciones informáticas que permitan contribuir al abordaje institucional de problemáticas de la población estudiantil de la UNAHUR.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Los objetivos y líneas de desarrollo del proyecto se trabajan principalmente en base a las disciplinas de desarrollo de software y de ciencia de datos que se focalizan en la construcción de un “Sistema de recomendaciones”.

Nos propusimos trabajar con un enfoque ágil, con resultados que se puedan compartir por fuera del equipo de desarrollo. Se busca utilizar un conjunto de lenguajes y tecnologías, que permitirá un despliegue sencillo de las aplicaciones construidas, una interfaz de usuario rica y agradable visualmente, y la posibilidad de que a futuro otras aplicaciones puedan utilizar los datos que se procesen y generen. Al mismo tiempo, la experiencia de construcción de software con herramientas del “estado del arte actual”. En particular, esta iniciativa fomenta la integración de la comunidad de Informática dentro del ámbito de la UNAHUR, lo que puede dar lugar a otros aportes.

La **ciencia de datos** aporta los conocimientos y técnicas necesarios para, por un lado, obtener la información necesaria sobre el comportamiento de la población estudiantil de la UNAHUR, y por otro proponer y evaluar alternativas para la caracterización de los estudiantes en riesgo de abandono, y el armado de propuestas de seguimiento.

En particular, nos proponemos realizar modelados utilizando técnicas de aprendizaje automático [*machine learning*], para comparar los resultados obtenidos con los de criterios basados en reglas [5 y 6]. Estas reglas se aplicarán a la selección de los estudiantes a quienes se envíen mensajes, y la determinación del contenido adecuado.

El proyecto también se nutrirá con aportes desde la ciencia de la educación [7]. La definición operativa de cuándo se considera a un estudiante en riesgo de abandono, y el cálculo de la inscripción propuesta, involucran un trabajo en conjunto entre especialistas en educación y de gestión académica.

El software en desarrollo se propone inicialmente como un trabajo multidisciplinariamente, atendiendo las miradas de distintos actores de la gestión de la universidad: especialistas de educación, de los coordinadores de carreras y de los institutos.

3. RESULTADOS OBTENIDOS/ESPERADOS

Este proyecto es la principal actividad que se desarrolla en el marco del CIDIA. Se lleva adelante el desarrollo del software “Sistema de recomendaciones”, con el objetivo de potenciar la retención de estudiantes de la UNAHUR. Los módulos centrales son:

- Configuración y envío de sugerencias de cursadas

- Monitoreo de estudiantes en riesgo de abandono y análisis de asistencia
- Predicción de abandono para alertas tempranas

Durante 2020 y 2021 el monitoreo de asistencia se vio postergado en medio de la pandemia donde la presencialidad perdió sentido.

Durante 2021 se vieron resultados en el desarrollo de las herramientas para potenciar la retención mediante envío de mensajes con sugerencias de cursadas.

El “Sistema de sugerencias” contempla los criterios de "estudiante en riesgo" y de "propuesta de inscripción" que decida la universidad, y luego llevar a cabo la implantación dentro del espacio informático de la institución, junto con la conexión con los sistemas de gestión académica y las plataformas de acompañamiento al aprendizaje con que cuente la misma.

Las definiciones conceptuales se realizaron a partir de consensos con diferentes actores de la gestión (coordinadores de carreras, secretaría académica, orientación estudiantil) para identificar las recomendaciones útiles para los estudiantes, considerando su carrera y trayectoria académica.

El “Sistema de recomendaciones” se trata de una tecnología para potenciar la comunicación de la Universidad con sus estudiantes, generando mensajes automatizados que puedan llegar a una gran cantidad de estudiantes, con contenidos personalizados. Se representa en un trabajo cíclico donde se aumentan las funcionalidades con el tiempo, aumentando las fuentes de información de base.

Se deben tener en cuenta los criterios y mecanismos de interacción con usuarios, para que la interfaz, de las aplicaciones informáticas a construir, resulte intuitiva y sencilla.

En 2021 se comenzó el desarrollo de un modelo de predicción de abandono, que considera variables de la historia académica disponible en la base de datos SIU-Guaraní de UNAHUR. La identificación de estudiantes en riesgo permitirá orientar acciones de intervención de manera temprana. Se realizará la primera prueba para predecir el abandono en 2022.

Las conclusiones se sumarán en otra etapa a las recomendaciones brindadas por mensaje a los estudiantes. En una etapa posterior se espera integrar también datos de la plataforma Moodle.

Uno de los resultados esperados es generar políticas que permitan disminuir el abandono estudiantil. Los resultados de este proyecto pueden contribuir a profundizar líneas de investigación aplicada sobre ciencia de datos para el análisis de la población estudiantil.

El proyecto resulta un punto de encuentro entre distintas áreas de la UNAHUR, e incluso con actores de otras universidades, siendo un vehículo para establecer lazos que permitan ulteriores iniciativas conjuntas.

Se han realizado varias actividades de difusión del proyecto en la UNAHUR y para ello se ha preparado un video que agiliza la su difusión y resume sus características principales (https://www.youtube.com/watch?v=ea_wXTBM9KE&t=1145s)

4. FORMACIÓN DE RECURSOS HUMANOS

La línea de investigación presentada colabora en la formación de varios estudiantes de la carrera Licenciatura en Informática a través de la modalidad de pasantía en el CIDIA, HUNAHUR. Participaron los siguientes estudiantes: Federico Pagliari, Cristian Saldivia, Jonathan Ybarra y Magalí Gaiani

5. BIBLIOGRAFÍA

- [1] Aiassa, H., Aliaga, J., Castellani, A., Daer, H., Glzman, M., Labaqui, F., ... & Torre, C. (2019). Reflexiones a 70 años de la gratuidad universitaria: políticas públicas y universidad.
- [2] Arias, M. F., Mihal, I., Lastra, K., & Gorostiaga, J. (2015). El problema de la equidad en las universidades del conurbano bonaerense en Argentina: un análisis de políticas institucionales para favorecer la retención. *Revista mexicana de investigación educativa*, 20(64), 47-69.
- [3] Chávez, M. J. (2020). Somos mujeres de sectores populares: ¿llegamos a la universidad?: aproximaciones al acceso de mujeres de sectores populares a las universidades públicas del conurbano bonaerense: el caso de la Universidad Nacional de Hurlingham (UNAHUR).
- [4] Mendonça, M. (2021). Una aproximación a las estrategias institucionales para lograr la permanencia de los estudiantes en las nuevas universidades del conurbano (2009-2016). *Espacios en Blanco. Revista de Educación*, 2(31), 275-286.
- [5] Hadley Wickham (2019), *Advanced R*, Second Edition (Chapman & Hall/CRC The R Series).
- [6] Brett Lantz (2018). *Machine Learning with R: Expert techniques for predictive modeling*.
- [7] Felie Munizaga, Maria Beatriz Cifuentes Orellana (2018). Retención y Abandono Estudiantil en la Educación superior Universitaria en América Latina y el Caribe: Una revisión Sistemática.

Educación a Distancia e Innovación Tecnológica

Autores: Claudia Russo¹, Mónica Sarobe², Benjamin Cicerchia³, Nicolás Alonso⁴, Gustavo Gnazzo⁵, Mariana Adó⁵, Marina Lilian Rodriguez⁵, Natalia Bendatti⁵, Valeria Cassera⁶, Bernabé Crespi⁵, Matías Contreras⁷, Gustavo Iglecias⁷, Tamara Ahmad⁸, Carla Decoud⁸, Mónica Tugnarelli⁸

{ccrusso, msarobe, lbcicerchia, nfallonso, ggnazzo, mado, nsbendati, celeste.nobl, bcrespi, vcasseras}@docentes.unsada.edu.ar, marina.rodriguez_2009@yahoo.com.ar, {mjcontrera, giclesias}@alumnos.unsada.edu.ar, tamaraahmad@unnoba.edu.ar, carladecoud@gmail.com, monica.tugnarelli@uner.edu.ar

Universidad Nacional de San Antonio de Areco, Buenos Aires, Argentina
Universidad Nacional del Noroeste de la Provincia de Buenos Aires, Instituto de Investigación y Transferencia- CIC

RESUMEN

Se presenta la continuidad de un proyecto de I+D+i (Investigación, Desarrollo e innovación) centrado no sólo en las tecnologías innovadoras y su relación con la Educación a Distancia (EaD) sino su relación con las distintas disciplinas. En el marco del presente y considerando que las tecnologías actuales plantean paradigmas que implican impacto directo en los modelos educativos, se investigará sobre las tecnologías informáticas aplicadas a diferentes contextos, haciendo énfasis en sus variantes y evoluciones en las diferentes disciplinas.

Se continuará indagando sobre distintas metodologías y técnicas que permitan dotar de interoperabilidad a los sistemas de información de la UNSaA, que tengan

relación con tecnologías planteadas en el presente proyecto, realizando experiencias concretas. Acompañará a la presente investigación la generación de una guía que posibilite la producción, administración y almacenamiento de contenidos digitales bajo la forma de Recursos Educativos Digitales (RED). La guía se propondrá desde una mirada interdisciplinaria que permita determinar y sistematizar aspectos determinantes en el desarrollo de materiales didácticos digitales.

En los próximos años las tecnologías innovadoras impactarán en el avance tecnológico, cobrando importancia los desarrollos que apliquen dicha tecnología, en el proyecto se trabajará principalmente con las siguientes tecnologías: Realidad Aumentada, EV3D, Realidad Virtual, Inteligencia Artificial, entre otras.

PALABRAS CLAVES

¹ Doctora en Ciencias Informáticas, Profesor Asociado CIC, UNNOBA, UNSADA

² Maestrando de la UNLP, Lic. en Informática, Profesor, UNNOBA, UNSADA

³ Doctorando de la UNLP, Profesor, UNSADA UNNOBA

⁴ Lic. en Informática, Profesor, UNNOBA, UNSADA

⁵ Lic. en Sistemas, Profesor, UNNOBA, UNSADA

⁶ Lic. en Informática, Profesor, UNSADA

⁷ Alumno avanzado de la Carrera Lic. en Informática, UNSADA

⁸ Investigador externo UNSADA

Entornos Virtuales, Innovación en Educación, Recursos Digitales, Realidad aumentada y virtual, EV3D, IA.

CONTEXTO

Las líneas de investigación a describir se enmarcan en el proyecto de investigación: Educación a Distancia e innovación Tecnológica, con lugar de trabajo en la Universidad Nacional de San Antonio de Areco presentado en la convocatoria de Subsidios a la Investigación 2022 ante la Secretaría de Investigación de la Universidad. Su objetivo es continuar con la investigación y la formación de recursos humanos sobre los aspectos tecnológicos del desarrollo de aplicaciones utilizando tecnologías innovadoras en diferentes contextos. El proyecto está integrado por un equipo interdisciplinario de investigadores, docentes y estudiantes pertenecientes a la Escuela de Desarrollo Productivo y Tecnológico, y se trabajará en conjunto con la Escuela de Desarrollo Social y Humano de la UNSAdA.

LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

En este período se trabajará con los fundamentos pedagógicos y didácticos incorporando a la educación a distancia y la educación presencial herramientas tecnológicas emergentes.

De esta forma se investigará cómo la informática impacta en el desarrollo de tecnologías innovadoras de manera de analizar, definir y desarrollar herramientas y estrategias innovadoras que impacten de manera responsable en el desarrollo de la educación y de la sociedad en general.

En los próximos años las tecnologías emergentes impactarán en el avance tecnológico y cobrarán importancia los desarrollos que apliquen dicha tecnología,

en el proyecto se trabajará principalmente con:

Realidad Aumentada (RA) que por definición, es el conjunto de tecnologías que permiten la superposición, en tiempo real, de imágenes, marcadores o información generados virtualmente, sobre imágenes del mundo real. En la realidad aumentada el mundo real y el mundo virtual se entremezclan para crear una realidad mixta en tiempo real. Es un recurso tecnológico que ofrece experiencias interactivas al usuario a partir de la combinación entre la dimensión virtual y la física, con la utilización de dispositivos digitales. La RA se aplica en contextos y campos diversos como la medicina, la publicidad, el campo militar, los dispositivos de navegación, la prospección hidrológica y geológica, agropecuario, ambiental o el mundo industrial entre otros.

Realidad virtual (RV) es un sistema tecnológico, basado en el empleo de distintos dispositivos, cuyo fin es producir una apariencia de realidad que permita al usuario tener la sensación de estar presente en ella. Su aplicación, aunque centrada inicialmente en el terreno de los videojuegos, se ha extendido a otros muchos campos, como la medicina, educación generando aulas virtuales o simulaciones de vuelo y laboratorios entre otros. La realidad virtual puede ser de dos tipos: inmersiva y no inmersiva. Los métodos inmersivos de realidad virtual con frecuencia se ligan a un ambiente tridimensional creado por una computadora, el cual se manipula a través de cascos, guantes u otros dispositivos que capturan la posición y rotación de diferentes partes del cuerpo humano. La realidad virtual no inmersiva también utiliza la computadora y se vale de medios como el que actualmente nos ofrece Internet, en el cual podemos interactuar en tiempo real con diferentes personas en espacios y ambientes que en realidad no existen sin la necesidad de dispositivos

adicionales a la computadora, ofreciendo un nuevo mundo a través de una ventana de escritorio.

Inteligencia artificial (IA) es la rama de las ciencias de la computación dedicada al desarrollo de agentes racionales no vivos, entendiendo como agente a cualquier cosa capaz de percibir su entorno, procesar tales percepciones y actuar en su entorno (proporcionar salidas). La inteligencia artificial convencional se basa en el análisis del comportamiento humano ante diferentes problemas. Los sistemas de IA actualmente son parte de la rutina en campos como educación, economía, medicina, ingeniería y la milicia, y se ha usado en gran variedad de aplicaciones de software, juegos de estrategia como ajedrez de computador y otros videojuegos. Otros ejemplos de sus usos se encuentran en el área de control de sistemas, planificación automática, la habilidad de responder a diagnósticos y a consultas de los consumidores, reconocimiento de escritura, reconocimiento del habla y reconocimiento de patrones.

Calidad de la educación en la virtualidad: ¿Qué significa una buena educación o calidad en el servicio en la Educación a Distancia? Debería significar lo mismo que para la educación presencial. La calidad de la educación institucional no depende de la modalidad educativa, sino de la calidad de los procesos que se viven y aprenden. ¿Por ello es importante definir indicadores para poder medir que tipo de educación estamos brindando?

Entornos virtuales 3D: con la tecnología tridimensional se pueden realizar diferentes herramientas que las personas u organizaciones pueden involucrar en su día a día para brindar un mejor servicio o para facilitar algunas de sus necesidades, las cuales son, pantallas 3D, juegos en 3D, impresiones 3D y entornos virtuales 3D. Los entornos virtuales 3D o espacios de simulación 3D son sistemas inmersivos, interactivos, personalizables, accesibles y

programables, que permiten diseñar actividades complejas.

RESULTADOS OBTENIDOS/ESPERADOS

Durante el proyecto, en la convocatoria anterior, se trabajó en el desarrollo de la base para la implementación de la educación digital en la UNSaDA, en este sentido se realizó un trabajo coordinado con el área Académica de la Universidad en la definición de protocolos para la implementación de la educación a distancia en el contexto de la pandemia.

Las medidas excepcionales dispuestas por la universidad ante la emergencia sanitaria COVID-19 impulsaron y potenciaron la puesta en funcionamiento del modelo de SIED presentado y aprobado por el Consejo Superior. Con anterioridad a la aprobación del SIED, por Resolución del Rector N° 78/2020, refrendada por Resolución 191/2020 del Consejo Superior, se aprobó el Protocolo de implementación del “Dispositivo de Trabajo de Educación Virtual por Causa de la Suspensión de Clases Presenciales - Seguimiento y Evaluación de las Actividades”. Si bien, como lo aclara explícitamente la Resolución 78/2020, dicho Protocolo viabiliza la aplicación de métodos y medidas de enseñanza virtual “sin que ello implique adoptar el mecanismo de Educación a Distancia como forma de enseñanza”, la situación en sí motivó e impulsó la definición y puesta en funcionamiento del SIED.

Por Resolución del Rector N° 78/2020 se dispuso un protocolo para el seguimiento y evaluación de las actividades virtuales desarrolladas durante la emergencia sanitaria. Desde la Secretaría Académica se realizó el seguimiento a través de tres sistemas informáticos: el correo institucional, SIU Guaraní y Moodle. Ante el surgimiento de potenciales

inconvenientes dispuso la intermediación de la Jefatura del Departamento Técnico, si se tratase de requerimiento o asesoramiento técnico-informático, y de la Coordinación de Bienestar Estudiantil, si se tratase de inconvenientes particulares planteados por alumnos.

Con anterioridad a la aprobación del SIED, por Resolución del Rector N° 78/2020, refrendada por Resolución 191/2020 del Consejo Superior, se aprobó el Protocolo de implementación del “Dispositivo de Trabajo de Educación Virtual por Causa de la Suspensión de Clases Presenciales - Seguimiento y Evaluación de las Actividades”. Si bien, como lo aclara explícitamente la Resolución 78/2020, dicho Protocolo viabiliza la aplicación de métodos y medidas de enseñanza virtual “sin que ello implique adoptar el mecanismo de Educación a Distancia como forma de enseñanza”, la situación en sí motivó e impulsó la definición y puesta en funcionamiento del SIED.

Tal como se viene mencionando anteriormente se realizó la presentación del Sistema Institucional a Distancia de la Universidad ante CONEAU obteniendo la aprobación por parte del Ministerio.

Se construyó, mediante herramientas de software de código abierto, un Entorno Virtual de Enseñanza y Aprendizaje (EVEA) que integra los sistemas de información para lograr una administración y gestión académica de forma centralizada que permite crear, administrar y llevar a cabo aulas virtuales, herramientas de aprendizaje, comunicación y colaboración.

Se presentó una Becas EVC – CIN Convocatoria 2019 a cargo del alumno Matías Contreras. Durante el desarrollo de la misma, se relevaron estudios de las diferentes propuestas para el desarrollo del EVEA se vislumbraron las ventajas de utilizar un entorno basado en un LMS de Software Libre como lo es Moodle, alojado sobre una infraestructura cloud (en la nube)

teniendo de manifiesto el contexto de la UNSAdA y su desarrollo tecnológico actual. La sólida arquitectura tecnológica con que cuenta la UNSAdA otorga sustentabilidad y facilita la proyección de líneas estratégicas para el desarrollo del SIED en el corto, mediano y largo plazo.

Se desarrolló una tesis de maestría en tecnología aplicada a la educación (UNLP) que actualmente se encuentra en evaluación para exposición.

Se continuará con la investigación y la formación de recursos humanos sobre los aspectos tecnológicos del desarrollo de aplicaciones utilizando tecnologías innovadoras en diferentes contextos.

FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está compuesto por docentes e investigadores formados y en formación pertenecientes a la Universidad Nacional de San Antonio de Areco, algunos de los cuales dirigieron becas CIN, y diversos trabajos finales de la carrera de Analista en Informática. Por otro lado, en agosto y septiembre de 2021 respectivamente, uno de los investigadores de este equipo defendió sus tesis de maestría obteniendo el título de Magíster en Tecnologías aplicadas a Educación otorgado por la Universidad Nacional de La Plata. Otros de los docentes investigadores del equipo se encuentran actualmente esperando fecha para la defensa de su tesis también en el Magíster en Tecnologías aplicadas a Educación otorgado por la Universidad Nacional de La Plata. Se espera en el 2022 tener dos presentaciones de Becas CIN de alumnos avanzados y una beca CIC.

BIBLIOGRAFÍA

Campi, W [et. al] (2018) ”La educación a distancia en Argentina a través de sus

normas: de la Ley 1597/1885 a la Resolución Ministerial 2641-E/2017” en Dari, N. y Bauman, P (Comps), Marcos regulatorios y modelos pedagógicos: un camino hacia la virtualización de la educación superior en el MERCOSUR, Bernal, Universidad Nacional de Quilmes.

Lezcano, H. y Suárez, M. (2021). “Educación Superior, tecnologías digitales y educación a distancia: impacto en la comunidad de estudiantes de la Universidad Nacional de San Antonio de Areco en relación a la aplicación de los dispositivos virtuales implementados en la enseñanza” Tesis de grado de la Lic. en Gestión Educativa, UNSaA.

López, S (2018) “Modelos pedagógicos en la educación a distancia: el caso de la Universidad Nacional de Quilmes en Argentina” en Dari, N. y Bauman, P (Comps), Marcos regulatorios y modelos pedagógicos: un camino hacia la virtualización de la educación superior en el MERCOSUR, Bernal, Universidad Nacional de Quilmes.

Pontoriero, F. A. (2021). E-learning en la educación superior argentina - Modelo de evaluación de calidad a partir del aporte de referentes clave Virtualidad, Educación y Ciencia, 22 (12), pp. 22-45.

Solla, L. M. (2019). “Integración de sistemas de información institucionales y entornos virtuales de aprendizaje: ¿necesidad operativa o un componente ausente en el SIED?” En el 8º Congreso internacional de la Red Universitaria de Educación a Distancia (RUEDA 2019).

AHMAD, T.; SAROBE, M.; RUSSO, C.; TESORE, J.; MORETTI, N. (2019) Experiencias Colaborativas en Entornos Virtuales 2D y 3D. Argentina. La Plata. 2019. Libro. Artículo Completo. Congreso. XIV CONGRESO NACIONAL DE TECNOLOGÍA EN EDUCACIÓN Y

EDUCACIÓN EN TECNOLOGÍA. Universidad Nacional de San Luis.

Baños González, N., Rodríguez García, T., Rajas Fernández, M. (2014): Mundos virtuales 3D para la comunicación e interacción en el momento educativo on-line. Historia y Comunicación Social. Vol. 19. Núm. 179 Especial Enero. Págs. 417-430. Recuperado de: <https://revistas.ucm.es/index.php/HICS/article/view/44967>.

Belloch, C. (2012) Entornos virtuales de aprendizaje. Unidad de Tecnología Educativa (UTE). Universidad de Valencia. Disponible en <https://www.uv.es/bellohc/pedagogia/EV A3.pdf>.

Cabero, J., & Llorente, M. (2007). La interacción en el aprendizaje en red: uso de herramientas, elementos de análisis y posibilidades educativas. RIED. Revista Iberoamericana de Educación a Distancia, 10(2), 97-123. Recuperado de: <https://doi.org/10.5944/ried.2.10.995>.

[13:30, 21/2/2022] Tamara: Campazzo, E. Guzmán, A. Martínez, M. Agüero, A. (2011). De la presencialidad a la interacción virtual 3D. Revista Calidad en la Educación Superior Programa de Autoevaluación Académica Universidad Estatal a Distancia. 2(1), 35 - 53. ISSN 1659-4703. Costa Rica. Recuperado de: <https://revistas.uned.ac.cr/index.php/revistacalidad/article/view/416>.

Cataldi, Z. Lage, F. J. Dominighini, C. (2013). Fundamentos para el uso de simulaciones en la enseñanza. Revista de Informática Educativa y Medios Audiovisuales. Vol. 10(17), págs.8-16, ISSN 1667-8338. Recuperado de: <http://laboratorios.fi.uba.ar/lie/Revista/Articulos/101017/A2mar2013.pdf>.

Entornos virtuales y herramientas digitales en escenarios educativos híbridos

Sanz Cecilia^{1,3}, Gorga Gladys¹, Gonzalez Alejandro¹, Zangara Alejandra¹, Iglesias Luciano¹, Ibáñez Eduardo¹, Violini Lucía^{1,2}, Fachal Adriana¹, Archuby Federico^{1,2}, Abásolo María José¹, Manresa-Yee Cristina⁴, Paula Dieser⁵, Pesado Patricia¹

¹Instituto de Investigación en Informática LIDI (III-LIDI). Centro Asociado CIC.
Facultad de Informática – Universidad Nacional de La Plata

²Becario/a UNLP

³Investigador Asociado de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

⁴Departamento de Ciencias Matemáticas e Informática. Universidad de las Islas Baleares, España

⁵Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa.

{csanz, ggorga, agonzalez, li, eibanez, lviolini, farchuby, mjabasolo, ppesado}@lidi.info.unlp.edu.ar, alejandra.zangara@gmail.com, afachal@hotmail.com, paula.dieser@gmail.com, cristina.manresa@uib.es

RESUMEN

Se presentan aquí algunas de las líneas de investigación y desarrollo del subproyecto “Metodologías y herramientas para la apropiación de tecnologías digitales en escenarios educativos híbridos”, correspondiente al Instituto de Investigación en Informática LIDI. En particular, se detallan los avances y logros alcanzados en este subproyecto en el año 2021. El trabajo se concentra en el desarrollo e innovación en el área de tecnologías digitales para escenarios educativos. Como parte de los objetivos se estudian técnicas, metodologías, y herramientas de diseño y desarrollo para diseñar y crear sistemas que puedan atender a los actuales contextos educativos; en los que se combinan diversas tecnologías, dispositivos, dinámicas de trabajo y personas con diferentes saberes y recorridos. Además, se participa fuertemente en la formación de recursos humanos en el área y en la cooperación con otras universidades del país y del exterior.

Palabras clave: materiales educativos digitales, entornos digitales para el aprendizaje, juegos serios, herramientas colaborativas, educación, educación especial

CONTEXTO

Este trabajo se enmarca en el subproyecto llamado “Metodologías y herramientas para la apropiación de tecnologías digitales en escenarios educativos híbridos” forma parte de un proyecto más general titulado: “Metodologías, técnicas y herramientas de Ingeniería de Software en escenarios híbridos. Mejora de proceso” (período 2018-2021), perteneciente al Instituto de Investigación en Informática LIDI, de la Facultad de

Informática de la Universidad Nacional de La Plata y acreditado por el Ministerio de Educación de la Nación. Participan en este proyecto docentes investigadores, tesis, becarios, y pasantes; se contribuye a la formación de recursos humanos y se coopera en estos temas tanto en el ámbito nacional como en el internacional.

1. INTRODUCCION

El subproyecto se organiza alrededor de diferentes ejes temáticos, que abarcan: el diseño y desarrollo de entornos digitales y materiales para mediar procesos educativos, metodologías y herramientas para el desarrollo y el seguimiento de actividades colaborativas, y el diseño y creación de juegos serios educativos. Como parte de las actividades del subproyecto, se aplican las investigaciones a diferentes niveles educativos y se investiga sobre sus efectos e impactos en estudiantes y docentes involucrados.

Los procesos educativos en la actualidad, y con mayor fuerza a partir de la situación de pandemia vivida, se encuentran atravesados por dispositivos, herramientas, redes sociales, y diversos entornos digitales (Kotilainen et al., 2020). Estas mediaciones conllevan el uso de datos para extraer información valiosa para los actores del escenario educativo así como técnicas de inteligencia artificial para ofrecer recorridos adaptados a los estudiantes según sus perfiles (Lion et al., 2020), entre muchos otros desafíos que involucran a las tecnologías, las personas y los procesos que los atraviesan (Spandre, Sanz & Dieser, 2021). Estos desafíos motivan algunas de las líneas que se llevan adelante en el marco de este sub-proyecto. Desde el área de la Informática, resulta de interés aportar al escenario educativo con herramientas que sean capaces de adaptarse a diferentes contextos, que transparenten datos de procesos que se realizan con ellas, y que

aporten al conocimiento y a la construcción colaborativa. Así uno de los ejes en los que se viene trabajando en este proyecto es el vinculado a los entornos digitales y materiales educativos, donde se están diseñando entornos educativos 3D para el trabajo con diferentes estudiantes (Fachal & Abásolo, 2021a; Quintín, Sanz & Zangara, 2016). Los entornos 3D propician la simulación a través de escenas recreadas, y propician la estimulación audiovisual (Quintín et al., 2016).

Al mismo tiempo, se investiga sobre el diseño de herramientas que aporten indicadores y datos de interés para procesos colaborativos en entornos virtuales de enseñanza y aprendizaje. Trabajos como el de Fidalgo-Blanco et al. (2012) ofrecen abrir el proceso en cada etapa e ir informando a los participantes acerca de su performance individual y grupal. En (Zangara & Sanz, 2020) se afirma que diversas formas de visualización de indicadores del proceso colaborativo, con representaciones adecuadas, pueden ayudar a los estudiantes en el desarrollo metacognitivo y, posteriormente, en la regulación de la actividad de colaboración.

En cuanto a la línea de juegos serios educativos, se abordan en el sub-proyecto tanto metodologías de diseño, como la creación y aplicación de juegos que se gestan con el fin de atender diversas necesidades del escenario educativo. Las metodologías de diseño de juegos, involucran atender a sus diversos componentes: los objetivos educativos y de entretenimiento, la historia, las mecánicas, la interfaz, entre otros (Archuby, Sanz & Manresa-Yee, 2020; Dörner et al., 2016). Profundizar en estos estudios es de interés para crear juegos que equilibren los objetivos pedagógicos y de entretenimiento y sean de utilidad para el desarrollo diversas competencias (Sandí Delgado et al., 2020). Se consideran estrategias de diseño co-participativo, centradas en las personas y se trabaja en metodologías ágiles para el diseño y desarrollo de las producciones que se llevan adelante en el sub-proyecto (Gros & Durall, 2020).

Las líneas de I+D+I que se abordan se vinculan con las temáticas de la Maestría y Especialización en Tecnología Informática Aplicada en Educación de la misma Facultad.

2. LINEAS DE INVESTIGACION / DESARROLLO/ INNOVACIÓN

Se presentan aquí las principales líneas de investigación, desarrollo e innovación abordadas en el marco del proyecto:

- Entornos digitales para la mediación de procesos educativos: entornos virtuales de enseñanza y aprendizaje, entornos 3D, redes sociales.

Funcionalidades y formas de acceso a estos espacios, trazabilidad de las actividades, estándares. Estrategias para el diseño e implementación de estos tipos de entornos.

- Materiales educativos digitales. Metodologías para su diseño y producción. Objetos de aprendizaje. Multimedia e hipermedia en escenarios educativos. Nuevos entramados de medios, soportes y lenguajes.
- Integración de TIC en procesos educativos. Hibridación de las modalidades educativas. Diseño de MOOC.
- Juegos Serios con diferentes paradigmas de interacción. Metodologías para su creación.
- Trabajo colaborativo mediado por TICs. Herramientas para la Autorregulación y el desarrollo de capacidades metacognitivas. Conceptualización, análisis y desarrollo de software y metodologías. Actividades colaborativas aprovechando dispositivos móviles.
- Formación de recursos humanos en el área de Informática. Pensamiento computacional, desarrollo de la empatía, y la resiliencia académica en estudiantes de carreras en Informática.

3. RESULTADOS OBTENIDOS

En esta sección se detallan los principales resultados alcanzados durante el 2021 e inicio de 2022.

Se continúa trabajando en el diseño y desarrollo del entorno IDEAS¹ (entorno virtual de enseñanza y aprendizaje - EVEA). En 2021 se desarrollaron mejoras en la herramienta de tareas que ofrece IDEAS, para ayudar a los participantes de un curso a distinguir el estado de las tareas (finalizada, nueva, pendiente, etc). Se llevó a cabo una reingeniería de la herramienta de autoevaluación, también atendiendo a sus diferentes estados, y enfatizando en mostrar la situación de cada estudiante respecto de la autoevaluación. Al mismo tiempo, se diseñaron nuevos reportes para atender a necesidades que surgieron durante la pandemia y ofrecer datos útiles a pedido de los docentes. En la actualidad IDEAS está siendo utilizado en diversos cursos de la Facultad de Informática y otras facultades de la UNLP. Se trabaja en el asesoramiento a docentes.

En el marco de estas temáticas en 2021 se finalizó un trabajo final de Especialización en Tecnología Informática Aplicada en Educación que indaga sobre el uso de estos entornos en la formación de usuarios de bibliotecas a nivel nacional (Kessler & Sanz, 2021). El trabajo está esperando su evaluación. También se desarrolló una tesis de maestría dirigida en el marco del proyecto, que aborda el diseño de una aplicación para la formación y entrenamiento de docentes en el uso de

¹ Proyecto IDEAS:

<https://proyectoideas.info.unlp.edu.ar/>

herramientas para entornos virtuales (Fernández, Gonzalez & Liberatore, 2021).

Se avanzó, al mismo tiempo, en una tesis de maestría sobre calidad de entornos virtuales de enseñanza y aprendizaje, dirigido por un miembro del proyecto (Ochoa Roblez, Esponda, Gorga & Astudillo, 2020).

En 2021 se profundizó en el diseño de un entorno 3D para el aprendizaje de personas con discapacidad auditiva (Fachal, Abásolo, Sanz, 2021a). Se crearon diferentes escenarios de trabajo en OpenSim, para llevar adelante una estrategia educativa con estudiantes con discapacidad auditiva. Al momento se desarrolló una prueba piloto de la primera etapa de esta estrategia, con 5 estudiantes y se están analizando resultados. Durante este año se dará continuidad a este trabajo completando el estudio de caso planificado. También se desarrolló una investigación sobre la creación de un diccionario de términos informáticos en el marco de una materia con estudiantes con discapacidad auditiva (Fachal, Sanz & Abásolo, 2021b). Hay una tesis vinculada a la creación de un simulador de entrevistas laborales que fue dirigida por un integrante del equipo y está en evaluación (Gonzalez et al., 2020). Otro resultado alcanzado en estas temáticas, se vinculó con el estudio de aplicaciones móviles para personas con discapacidad auditiva que resultó en una publicación (Herrera, Manersa-Yee, & Sanz, 2021). Este resultado forma parte de la agenda de cooperación con la Universidad de Islas Baleares y la Universidad Nacional de Santiago del Estero.

En una tesis doctoral de uno de los miembros del proyecto se está avanzando en el diseño y creación de MarCOA, un framework para la creación de objetos de aprendizaje, que sigue la metodología CROA (Violini, Sanz, & Pesado, 2021). Actualmente, se dirige una práctica profesional de un estudiante de Ingeniería en Computación que aborda parte de la implementación de este framework. Se espera finalizar la tesis en el 2022.

En 2021 se continuó con el proyecto Cap4city (Erasmus), en el que se aplicó la metodología de creación de MOOC elaborada en el equipo (Zangara & Sanz, 2020).

Al mismo tiempo, se destacan los estudios que se desarrollan, vinculados al sub-proyecto, en temas de autorregulación del aprendizaje en entornos mediados por tecnologías digitales (Dieser, Zangara & Sanz, 2019). En este sentido los resultados alcanzados pueden aportar al diseño de cursos online y al de entornos y herramientas que consideren los hallazgos encontrados. En 2021 se llevaron adelante dos charlas en el marco del plan de actividades del Centro de Innovación y Transferencia Tecnológica (CIyTT) que permitieron la discusión de estas temáticas entre los participantes, en su mayoría docentes de diferentes universidades del país (Dieser, Sanz & Zangara, 2021, 2020). Actualmente, se está iniciando una línea relacionada con la investigación sobre resiliencia académica y se profundizan los estudios realizados en el marco de un proyecto de innovación docente de la Universidad de Zaragoza en el

que se participa, que abordan el desarrollo de la competencia de empatía en estudiantes de carreras en Informática (Aguelo et al., 2020).

En cuanto a los temas vinculados a la mediación de procesos colaborativos, se está dirigiendo una tesis doctoral, que profundiza los aportes de la tesis doctoral de uno de los miembros del equipo (Zangara & Sanz, 2020). Se propone la creación un *dashboard* que se integra a Moodle que permite implementar *mirroring* de procesos colaborativos (Vazquez-Bermudez et al., 2021; van Leeuwen & Rummel, 2020). La tesis actualmente está avanzando en la implementación. También se dirige una tesis de maestría donde bajo el modelo de Comunidad de Indagación, se estudian las interacciones entre los actores de procesos educativos mediados por tecnologías digitales (Spandre, Sanz & Dieser, 2021).

Relativo a los temas de juegos serios educativos, durante 2021, se encaró un proyecto financiado por la Facultad de Informática en el que se buscó investigar la facilidad de uso de la metodología DIJS para la creación de estos juegos. Parte de la metodología DIJS se aplicó en el diseño de dos juegos: Escapados (juego en un entorno 3D con interacción tangible para abordar temas de Química y Matemática), y EmpoderAR: un juego de realidad aumentada para dar a conocer a algunas de las mujeres destacadas en la historia de la Informática. La metodología DIJS es uno de los resultados alcanzados en el marco de una tesis de maestría de un miembro del proyecto (Archuby, Sanz, & Pesado, 2020; Archuby, Sanz, & Manresa-Yee, 2020). Se avanzó también con la investigación en cooperación con la Universidad Nacional de Río Negro, donde se dirige un proyecto vinculado a esta temática. Como parte de las actividades desarrolladas en 2021 se aplicaron juegos serios en diversos contextos educativos (Livos et al., 2021). Además, se está dirigiendo un trabajo de fin de carrera sobre el tema ExerGames. Los ExerGames se utilizan con el objetivo de fomentar la actividad física de los jugadores. En el trabajo se desarrolló un mando para sensar los movimientos del jugador y contabilizarlos como parte de la dinámica de los juegos que integren dicho mando (Del Gener, Sanz & Iglesias, 2021).

En 2021 se aplicaron los juegos serios educativos creados en el marco del proyecto: Innovática (juego de Realidad Virtual sobre Innovadores de la Informática); Albores que es un juego en una sala interactiva y que fue aplicado en grupos de prueba piloto y se continúa desarrollando; HuVi Parque Nacional Iguazú y HuVi Ischigualasto Talampaya que se presentaron en charlas del CIyTT y del Programa Nexos durante el 2021. Además, se han creado nuevos juegos que serán aplicados y cuyo impacto se evaluará durante 2022.

En cuanto a los proyectos vinculados con la temática y los acuerdos de cooperación, el III- LIDI participa en los siguientes:

- Se cuenta con un acuerdo de colaboración en estos temas con la Universidad de Zaragoza y se trabaja en forma conjunta.
- Se tienen un acuerdo de cooperación con la Universidad de Islas Baleares, en particular se trabaja con el Departamento de Ciencias Matemáticas e Informática. En este contexto se dirigen tesis y se participa de proyectos conjuntos.
- Se finalizó la participación en el proyecto REFORTICCA: Recursos para el Empoderamiento de FORMadores en TIC, Ciencias y Ambiente. Proyecto que se desarrolla en el marco de los Proyectos de Innovación y Transferencia en Áreas Prioritarias de la Pcia. de Buenos Aires (PIT-AP-BA).
- Se participa del Proyecto Nexos de articulación entre la escuela y la universidad, a partir del cual se llevan a cabo acciones con diferentes escuelas de la región de La Plata y se participa de ferias y exposiciones de ciencia, tecnología y educación.
- Se coopera con la Universidad Nacional de Santiago del Estero y con la Universidad Nacional de Río Negro con asesoría en el primer caso y dirección de un proyecto en el segundo, en vinculación a los ejes presentados aquí.
- Se participa en el proyecto PERGAMEX y en particular en el subproyecto RTI2018-096986-B-C31: “Design of pervasive gaming experiences for intergenerational social and emotional well-being (PERGAMEX-INTERGEM)”.
- Se participa en proyectos de Innovación Docente de la Universidad de Zaragoza.

4. FORMACION DE RECURSOS HUMANOS

En estas líneas del subproyecto se realiza formación de recursos humanos a través de la dirección de becas, tesis de doctorado, maestría y trabajos finales de especialización y tesinas de grado.

En 2021, se cuenta como resultado con 2 trabajos de tesis finalizados diciembre de 2020 y cuyos libros de tesis fueron publicados en 2021. Además, de un trabajo de tesis de maestría dirigido por uno de los miembros del equipo entregado para su evaluación que será defendido en marzo de 2022. Se están dirigiendo dos tesis de maestría que avanzan sobre los temas de autorregulación del aprendizaje y comunidad de indagación para la interacción entre los participantes de procesos educativos mediados por tecnologías digitales, respectivamente. Se finalizó además un trabajo final de especialización, en temas vinculados a entornos virtuales, que está esperando su evaluación. En la línea de entornos virtuales 3D se espera concluir con una tesis doctoral.

5. BIBLIOGRAFIA

- Aguelo, A., Sanz, C., Coma, T., Baldassarri, S. & Álvarez, P. (2020). Educational Activity Proposal For The Development Of Empathy In Beginner Students Of Computer Science, EDULEARN20 Proceedings, pp. 4989-4995.
- Archuby, F., Sanz, C. & Manresa-Yee, C. (2020). Metodologías de diseño y desarrollo para la creación de juegos serios digitales. Tesis de maestría finalizada. Sep. 2020. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/111123>
- Archuby, F., Sanz, C. & Pesado, P. (2020). Experience Analysis for the Use of Desafiate Serious Game for the Self-assessment of Students. In: Pesado P., Arroyo M. (eds) Computer Science – CACIC 2019. CACIC 2019. Communications in Computer and Information Science, vol 1184. Springer, Cham. https://doi.org/10.1007/978-3-030-48325-8_8
- Astudillo, G., Sanz, C. & Santacruz-Valencia, L. (2017). Proceedings of the Twelfth Latin American Conference on Learning Technologies (LACLO), 2017, ISBN: 978-1-5386-2376-3, págs. 1-4, doi. 10.1109/LACLO.2017.8120939.
- Carvalho, L.; Yeoman, P. (2021) Performativity of Materials in Learning: The Learning-Whole in Action. Journal of New Approaches in Educational Research, [S.l.], v. 10, n. 1, p. 28-42, jan. 2021. ISSN 2254-7339. Available at: <https://naerjournal.ua.es/article/view/627>
- Del Gener, A., Iglesias, L. & Sanz, C. (2021). Exergames: propuesta de un gamepad para sensor movimientos del jugador. Tesina de Grado. Propuesta aprobada. En realización.
- Dieser, P., Sanz, C., Zangara, A. (2021). Charla Autorregulación del aprendizaje en entornos mediados por tecnologías digitales. Plan de actividades del CIyTT.
- Dieser, P., Sanz, C., Zangara, A. (2020). Capítulo: Autorregulación del aprendizaje en contextos educativos mediados por tecnologías digitales. Teoría, investigación y aplicaciones en la educación superior iberoamericana. Libro: Aprendizaje y tecnologías: habilidades del presente, proyecciones de futuro. Editorial Noveduc. Julio 2020. 978-987-538-764-5
- Dörner, R., Effelsberg, W., Göbel, S., Wiemeyer, J. (2016). Serious Games. Foundations, concepts and practice. Springer. Alemania
- Fachal, A. & Abásolo, M.J. (2021a). Entorno Virtual 3D en OpenSim para el trabajo con estudiantes con discapacidad auditiva. Demo presentada en el Congreso TE&ET 2021.
- Fachal, A. S., Abásolo, M. J., & Sanz, C. V. (2021b). Dictionary of Computer Terms in LSA with Operational Signs proposed by and for Hearing-Impaired Students. IEEE Revista Iberoamericana de Tecnologías del Aprendizaje.

- Fachal, A. S., Abásolo, M. J., & Sanz, C. V. (2019). Experiences in the Use of ICT and Digital Ramps for Students in Tertiary Education with Visual or Hearing Impairment. In Argentine Congress of Computer Science (pp. 369-388). Springer, Cham.
- Fidalgo-Blanco, A., Lerís, D., Sein-Echaluce, and F. J. García-Peñalvo, M.L. (2013). Indicadores para el seguimiento y evaluación de la competencia de trabajo en equipo a través del método CTMTC, [Online] Available in: <https://gredos.usal.es/handle/10366/122531>
- Fernández, G. V., González, A. H., & Liberatore, G. (2021). Diseño de una APP Web para la formación y el entrenamiento de docentes en TIC: el caso de las carreras BIBES y LICAD, UNMdP. In XVI Congreso de Tecnología en Educación & Educación en Tecnología-TE&ET 2021 (La Plata, 10 y 11 de junio de 2021).
- González, A. H., Quintana, N., Vallejo, A. E., & Pereyra, J. M. (2020). Capacitación de adultos hipoacúsicos mediante un entorno multimedia basado en la simulación de una entrevista laboral. In XV Congreso Nacional de Tecnología en Educación y Educación en Tecnología (TE&ET 2020). <http://sedici.unlp.edu.ar/handle/10915/103672>
- Gros, B., & Durall, E. (2020). Retos y oportunidades del diseño participativo en tecnología educativa. *EduTec. Revista Electrónica De Tecnología Educativa*, (74), 12-24. <https://doi.org/10.21556/edutec.2020.74.1761>
- Herrera, S. I., Manresa-Yee, C., & Sanz, C. V. (2021). Mobile learning for hearing-impaired children: Review and analysis. *Universal Access in the Information Society*, 1-19.
- Kotilainen, S., Okkonen, J., Vuorio, J. & Leisti, K. (2020). Youth Media Education in the Age of Algorithm-Driven Social Media. En D. Frau-Meigs, S. Kotilainen, M. Pathak-Shelat, M. Hoehsmann & S. R. Poyntz (Eds.), *The Handbook of Media Education Research* (pp. 131–139). Hoboken, NJ: John Wiley & Sons, Ltd.
- Lion, C. (2020). Aprendizaje y tecnologías: habilidades del presente, proyecciones de futuro. *Noveduc Gestión*. ISBN 978-987-538-756-0
- Lovos, E., Sanz, C., Goin, M., Ricca, M., Molina, C., Gil, E., Basciano, I., Gastaminza, M. (2020). *Juegos Serios Móviles. Diseño, Desarrollo e Integración En Escenarios Educativos*. Workshop de Investigadores en Ciencias de la Computación. WICC2020
- Ochoa Roblez, J., Esponda, S. & Astudillo, G. (2020). Análisis del estado del arte de los modelos de calidad de Entornos Virtuales de Enseñanza y Aprendizaje. Trabajo final de Especialización en TIAE. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/89019>
- Ochoa Roblez, J., Esponda, S., Gorga, G. & Astudillo, G. Diseño de un modelo de evaluación de calidad de entornos virtuales de enseñanza aprendizaje basado en la familia de calidad ISO/IEC 25000. Propuesta de tesis de Maestría en TIAE.
- Quintín, E., Sanz, C., & Zangara, A. (2016, October). The impact of role-playing games through Second Life on the oral practice of linguistic and discursive sub-competences in English. In 2016 International Conference on Collaboration Technologies and Systems (CTS) (pp. 148-155). IEEE.
- van Leeuwen, A. & Rummel, N. (2020) Comparing teachers' use of Mirroring and advising dashboards. In: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, Frankfurt Germany, pp. 26–34, March 2020. <https://doi.org/10.1145/3375462.3375471>
- Vázquez-Bermúdez, M., Sanz, C., Zangara, M. A., & Hidalgo, J. (2021). Visualization Tools for Collaborative Systems: A Systematic Review. In *International Conference on Technologies and Innovation* (pp. 107-122). Springer, Cham.
- Violini, L.; Sanz, C. & Pesado, P. (2020) metodologías y framework para el diseño y la creación de objetos de aprendizaje. *Revista Investigación joven* (ISSN 2314-3991), vol. 6, num. Especial, págs. 176-177, 2019.
- Zangara, A. & Sanz, C. (2020) Trabajo colaborativo mediado por tecnología informática en espacios educativos. Metodología de seguimiento y su validación. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*, no. 25, pp. 8-20, 2020. doi: 10.24215/18509959.25.e1
- Zangara, A. & Sanz, C. (2019). Del trabajo grupal al colaborativo. antecedentes, conceptualización y propuesta de abordaje didáctico. *Signos Universitarios*, (54). Referato Nacional.

Diseño de una aplicación para la formación y entrenamiento de docentes en el uso de herramientas para entornos virtuales (EVEA): el caso de las carreras a distancia del Departamento de Ciencia de la Información de la Universidad Nacional de Mar del Plata.

Tesis de Maestría en Tecnología Informática Aplicada en Educación, Facultad de Informática, Universidad Nacional de La Plata

Tesista: Fernández, Gladys Vanesa
gvfernan07@gmail.com

Director: Alejandro Héctor González (III LIDI, Facultad de Informática, UNLP). **Codirector:** Gustavo Liberatore (Departamento de Ciencia de la Información, Facultad de Humanidades, Universidad Nacional de Mar del Plata)

Fecha de exposición: noviembre de 2021

MOTIVACIÓN

La motivación para realizar esta propuesta de tesis surge de la necesidad de una formación docente más profunda en herramientas tecnológicas y recursos para educación a distancia, sumado a que, como docente y autora central de este proceso y de los desarrollos realizados, he transitado gran parte de la historia de las carreras a distancia.

En el caso particular de las carreras, Bibliotecario Escolar y Licenciatura en Bibliotecología y Documentación, la intervención pedagógica de los docentes ha consistido en transmitir a los alumnos las claves de lectura de los materiales didácticos y de la bibliografía obligatoria, según la secuencia de un Plan de Trabajo Docente (PTD), presentado al iniciar la asignatura. Los procesos educativos en las carreras a distancia han sido desarrollados por la mayoría de los docentes en términos de la poca utilización de aplicaciones informáticas que potencien las capacidades de intermediación tecnológica en la apropiación y discusión de los contenidos ofrecidos. La intervención tecnológica de los docentes de estas carreras se ha limitado a la utilización de las herramientas básicas del EVEA (entorno virtual de enseñanza y aprendizaje), y en muy pocos casos a la innovación en el uso de recursos provenientes del propio software o de fuentes externas. El EVEA, en este sentido, se ha comportado como un “contenedor” administrado por prácticas docentes presenciales cuando debería ser un espacio donde aplicar estrategias didácticas adaptadas a un entorno tecnológico.

OBJETIVOS

- Analizar y diagnosticar las competencias tecnológicas y digitales de los docentes de las carreras: Bibliotecario Escolar y Licenciatura en Bibliotecología y Documentación a Distancia.
- Crear una aplicación web para el entrenamiento y formación de los docentes de las carreras: Bibliotecario Escolar y Licenciatura en Bibliotecología y Documentación a distancia con la finalidad de que se realice un aprovechamiento de los recursos provistos por el EVEA Moodle y las herramientas externas al software.

Objetivos particulares

- Relevar y analizar la situación actual del cuerpo docente en cuanto a competencias digitales y tecnológicas.
- Identificar recursos y actividades dentro del EVEA Moodle para el entrenamiento y la formación de los docentes.
- Evaluar, valorar y seleccionar herramientas complementarias en la Web para la actividad de enseñanza y aprendizaje a distancia.
- Generar pautas para el entrenamiento y formación de los docentes en los recursos y las herramientas definidas anteriormente a través de material multimedia y las estrategias de simulación.
- Crear una aplicación Web para la formación y el entrenamiento de los docentes en recursos, actividades y herramientas en el contexto de Moodle mediante la utilización de material multimedia y estrategias de simulación.

APORTES

Este trabajo presenta el desarrollo de una herramienta para la formación y el entrenamiento en tecnologías de educación a distancia, destinada a los docentes de las carreras Bibliotecario Escolar y Licenciatura en Bibliotecología y Documentación del Departamento de Ciencia de la Información, Facultad de Humanidades, Universidad Nacional de Mar del Plata.

La tesis consta de dos partes. En la primera parte se desarrolla la detección formal de la necesidad de instancias de formación en competencias tecnológicas, realizada a través del relevamiento del estado actual. La segunda parte, toma los resultados generados por el relevamiento para la confección de una aplicación para la formación de los docentes, teniendo en cuenta los siguientes aspectos:

- los recursos que posee el Entorno Virtual de Enseñanza y Aprendizaje que funciona actualmente para las carreras a distancia;
- los recursos disponibles, por fuera del EVEA, para la utilización en diferentes • escenarios del proceso de enseñanza y aprendizaje en educación a distancia;
- un plan de formación virtual y autónoma para cada recurso.

El trabajo de tesis pone énfasis en los siguientes puntos:

- La importancia de la transversalidad de las competencias digitales y tecnológicas.
- La incorporación de competencias tecnológicas y digitales no solo para el proceso de enseñanza y aprendizaje sino también para las tareas de gestión e investigación propias del perfil docente universitario.
- Un sostenimiento de las políticas de formación a lo largo del tiempo, teniendo en cuenta la dinámica y la evolución de las TIC.

Por último, es definida la metodología de testeo y evaluación para el producto final. A través de este trabajo fue posible evidenciar la necesidad de instancias de formación para los docentes de las carreras a distancia. Si bien la encuesta arrojó como resultado que el profesorado tiene un uso limitado de las tecnologías, pudo ponerse en evidencia la necesidad de un marco de condiciones pertinentes para que conozcan, amplíen y generen un uso innovador de la tecnología en la práctica docente a distancia.

Con estos resultados de la encuesta se pudieron detectar los puntos específicos con carencias de los docentes de BIBES y LICAD y arribar a la conclusión de que el profesorado de las carreras a distancia deberá:

- Aprovechar el uso de herramientas colaborativas en el proceso de enseñanza y aprendizaje.
- Contemplar la inclusión digital bajo el diseño universal de materiales educativos.
- Incorporar herramientas para la dinamización de las tutorías en cada asignatura.
- Complementar el uso del EVEA con otros ambientes digitales: aplicaciones semánticas, herramientas de gamificación, realidad aumentada. etc

LÍNEAS DE TRABAJO FUTURO

Como líneas de trabajo futuro, se plantea la realización de un nuevo análisis del estado de las competencias tecnológicas y digitales, luego de la implementación de la aplicación Web como política de formación de los docentes de las carreras BIBES y LICAD con el objetivo de analizar el grado de evolución en el manejo y la implementación de TIC en el perfil docente universitario.

Con respecto a la aplicación Web el trabajo a futuro se desarrollará:

- La posibilidad de la carga colaborativa de herramientas y recursos por parte de otros formadores.
- La utilización de la aplicación Web en otras dependencias y unidades académicas.
- La implementación de la aplicación Web dentro de un plan de formación para docentes en TIC por parte del SIED-UNMdP.

Herramientas de software para soporte en la enseñanza y aprendizaje de la arquitectura x86

Marcelo A. Colombani, José M. Ruiz, Amalia G. Delduca, Marcelo A. Falappa

Universidad Nacional de Entre Ríos, Av. Tavella 1424, (E3202KAC), Concordia, Entre Ríos, Argentina
Universidad Nacional del Sur Av. Alem 1253, (B8000CPB) Bahía Blanca, Argentina

[marcelo.colombani](mailto:marcelo.colombani@uner.edu.ar), [jose.ruiz](mailto:jose.ruiz@uner.edu.ar), [amalia.delduca](mailto:amalia.delduca@uner.edu.ar)}@uner.edu.ar, mfalappa@cs.uns.edu.ar

RESUMEN

Existe un consenso creciente en el uso de herramientas de simulación en la enseñanza para procesos dinámicos complejos, como las operaciones intrínseca de la computadora, permiten representar de forma visual e interactiva la organización y arquitectura interna de la computadora, facilitando así la comprensión de su funcionamiento por parte de los alumnos y el desarrollo de los temas por parte del docente. En este contexto, los simuladores juegan una pieza clave en el campo de la Arquitectura de Computadores, permitiendo conectar fundamentos teóricos con la experiencia práctica simplificando abstracciones y haciendo más rica la labor docente.

La arquitectura x86 es ampliamente utilizada en computadoras de escritorio y servidores. Este documento pretende realizar una comparación de los simuladores x86 que más se adecuan en el dictado de la asignatura Arquitectura de Computadoras de la carrera Licenciatura en Sistemas, establecer los criterios de evaluación y evaluar los simuladores seleccionados de acuerdo con estos criterios.

Palabras clave: x86, x86-64, simulador, aprendizaje, enseñanza, arquitectura de computadoras.

CONTEXTO

La presente investigación surge en el marco del proyecto de investigación I/D novel **PID-UNER 7065**: “Enseñanza/aprendizaje de asignatura Arquitectura de Computadoras con herramientas de simulación de sistemas de cómputos”. El Proyecto es llevado a cabo en la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos, se vincula directamente con la asignatura Arquitectura en Computadoras que se dicta en segundo año de la carrera Licenciatura en Sistemas perteneciente a la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos.

Dentro del marco de la presente línea de investigación se está desarrollando una tesis de maestría, en la carrera Maestría en Sistemas de Información dictada en la Facultad de Ciencias de la Administración, UNER en el área de ARSO.

1. INTRODUCCIÓN

Simulación

La simulación es un término de uso diario en muchos contextos: medicina, militar, entretenimiento, educación, etc., debido a que permite ayudar a comprender cómo funciona un sistema, responder preguntas como "qué pasaría si", con el fin de brindar hipótesis sobre cómo o por qué ocurren ciertos fenómenos.

Para continuar, se define simulación como el

proceso de imitar el funcionamiento de un sistema a medida que avanza en el tiempo. Entonces para llevar a cabo una simulación, es necesario desarrollar previamente un modelo conceptual que representa las características o comportamientos del sistema, mientras que la simulación representa la evolución del modelo a medida que avanza en el tiempo. [1,3].

Con los avances en el mundo digital, la simulación se ha convertido en una metodología de solución de problemas indispensable para ingenieros, docentes, diseñadores y gerentes. La complejidad intrínseca de los sistemas informáticos los hace difícil comprender y costosos de desarrollar sin utilizar simulación [3].

Aporte pedagógico

Muchas veces en el ámbito educativo, resulta difícil transmitir fundamentos teóricos de la organización y arquitectura interna de las computadoras debido a la complejidad de los procesos involucrados. Si sólo incorporamos los medios de enseñanza tradicionales, como puede ser una pizarra, un libro de texto o diapositivas, los mismos tienen una capacidad limitada para representar estos fundamentos. En consecuencia, es imprescindible un alto nivel de abstracción por parte del alumno para desarrollar un modelo mental adecuado para capturar la organización y arquitectura interna de las computadoras [4],[5].

Es evidente la necesidad de utilizar nuevas tecnologías como recurso didáctico y como medio para la transferencia de conocimiento, ya que resultan de gran ayuda para que los alumnos relacionan conceptos abstractos con reales, permite situar al alumno en un contexto que imite algún aspecto de la realidad; en ese ambiente, el alumno podrá detectar problemáticas similares a las que podrían producirse en la realidad, logrando un mejor

entendimiento por medio del trabajo exploratorio, inferencia, aprendizaje por descubrimiento y desarrollo de habilidades [6,7].

Simuladores

Un simulador de arquitectura es un software que imita el hardware de un sistema. El simulador se centra principalmente en la representación de los aspectos arquitectónicos y funciones del hardware simulado. El uso de herramientas de simulación permite realizar cambios, pruebas y ejecución de programas sin temor de dañar ningún componente o por falta de la computadora [8].

Algunos softwares ofrecen una representación en forma visual e interactiva de la organización y arquitectura interna de la computadora, facilitando así la comprensión de su funcionamiento, como ser los simuladores Assembly debugger (x86), Simple 8-bit Assembler Simulator, Microprocessor Simulator, Simulador de ensamblador de 16 bits y Emu8086. En este sentido, los simuladores juegan una pieza clave en el campo de la Arquitectura de Computadores, permitiendo conectar fundamentos teóricos con la experiencia práctica, simplificando abstracciones y facilitando la labor docente [9,13].

Repertorio de instrucciones x86

El repertorio de instrucciones de la arquitectura x86 es la más utilizada en computadoras de escritorio y servidores del mundo. Inició con el procesador Intel 8086 en el año 1978 como arquitectura de 16 bits. Después evolucionó hasta una arquitectura de 32 bits cuando apareció el procesador Intel 80386 en el año 1985, denominada i386 o x86-32. AMD amplió esta arquitectura de 32 bits a una de 64 bits. Intel adoptó las extensiones de la arquitectura de AMD de 64 bits, también denominada AMD64 o Intel 64

[14,16].

Un procesador x86-64 mantiene la compatibilidad con los modos x86 existentes de 16 y 32 bits, y permite ejecutar aplicaciones de 16 y 32 bits, como así también de 64 bits. Esta compatibilidad hacia atrás protege las principales inversiones en aplicaciones y sistemas operativos desarrollados para la arquitectura x86 [14,16].

Por ello, la enseñanza de la arquitectura x86 es de gran relevancia en la asignatura Arquitecturas de Computadoras debido a los diferentes temas que aborda.

Asignatura Arquitectura de Computadoras

Los alumnos de la asignatura Arquitectura de Computadoras no solo deben conocer la estructura y el funcionamiento interno de la computadora, sino que, idealmente, deben tener una experiencia práctica activa con dicha arquitectura.

Para proporcionar esta experiencia es necesario un laboratorio con el hardware necesario y el tiempo para que los alumnos se vuelvan competentes en el uso de herramientas para trabajar con el hardware. Por este motivo, muchos simuladores han sido desarrollados, ayudando al alumno a comprender el funcionamiento y la estructura del computador proporcionando valiosas experiencias de aprendizaje [17].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Si bien ya existe un estudio comparativo acerca de los simuladores de la arquitectura x86, el mismo está hecho bajo los siguientes criterios: velocidad, modelos de simulación y error experimental [18]. Sin embargo, desde este trabajo se propone un enfoque diferente, evaluar los simuladores x86 bajo otros criterios de evaluación vinculados a sus

características de simulación y a determinados contenidos abordados en la asignatura Arquitectura de Computadoras de la Carrera de Licenciatura en Sistemas de la Universidad Nacional de Entre Ríos.

Cabe destacar que existen estudios del año 2009 [9] y del año 2012 [10] que evalúan diferentes simuladores para abordar diferentes temas en el dictado de los cursos de Arquitectura de Computadoras, en general estos estudios evalúan simuladores en términos de dos categorías predefinidas: una referida a las características de la simulación, como ser granularidad, usabilidad, disponibilidad, presentación visual, flujo de simulación, etc., y otra sobre la cobertura de los contenidos preestablecidos en las currículas.

La diferencia de este trabajo respecto a los mencionados anteriormente [9,10] radica en evaluar los simuladores x86 en relación a una temática específica de la currícula como es el repertorio de instrucciones, funcionamiento de la CPU, ciclo de búsqueda y ejecución. En cuanto a las características de simulación se pretende incorporar nuevos criterios de evaluación (última actualización, documentación, editor para escribir programas en Assembler, adaptación a las extensiones de la arquitectura de 32 y 64 bits, etc).

En base a lo expuesto, se buscará definir criterios de evaluación y comparar los simuladores x86 y x86-64 disponibles según los criterios de evaluación preestablecidos, bajo dos categorías: una referida a las características de la simulación que analizan la funcionalidad de los simuladores seleccionados y otro referido a contenidos específicos de la asignatura Arquitectura de Computadoras que cubren los simuladores seleccionados.

3. RESULTADOS OBTENIDOS/ESPERADOS

Este proyecto de investigación se encuentra en su fase de iniciación y, entre los objetivos buscados podemos remarcar:

- Estudiar y evaluar diferentes herramientas de simulación de procesadores de la arquitectura x86.
- Definir criterios de evaluación de las herramientas de simulación para su utilización en el dictado de clases.
- Comparar diferentes herramientas de simulación según los criterios de evaluación preestablecidos.
- Analizar el impacto de la jerarquía de memorias en la ejecución de un programa utilizando las herramientas de simulación de procesadores.
- Generalizar dichos conceptos a otras arquitecturas, como por ejemplo las actuales basadas en 64 bits.

Se avanzó sobre el objetivo de la definición de criterios de evaluación, permitiendo de esta manera comenzar a comparar las diferentes herramientas de simulación.

Se han definido 7 criterios de evaluación:

1. **Usabilidad:** se refiere a la capacidad de un software de ser comprendido, aprendido, usado y ser atractivo para el usuario, en condiciones específicas de uso. Escala usabilidad (difícil-media-fácil).
2. **Editor:** soporte para escribir código fuente en lenguaje ensamblador. Escala editor (baja-media-alta).
3. **Documentación:** disponibilidad de soporte para el aprendizaje, repertorio de instrucciones, manual de usuario. Escala documentación (mínima-media-completa)
4. **Ejecución de simulación:** facilidad para controlar la simulación. Escala ejecución de simulación (baja-media-alta).

5. **Nivel de especificación de la Organización y Arquitectura del sistema simulado:** nivel de implementación del set de instrucciones, memoria, módulos de E/S, etc. Escala nivel de especificación x86 (mínima-media-completa).

6. **Características del desarrollo del producto software:** tipo de licencia open source o privativas, fecha de última versión, Web/Escritorio, uso académico. Escala producto software (mala-buena-muy buena).

7. **Cobertura de los contenidos preestablecidos en las currícula:** se busca que la herramienta abarque o se ajuste a la mayor cantidad de tópicos de la asignatura Arquitectura de Computadoras, escala cobertura contenido (baja-media-alta):

- Estructura y funcionamiento de la computadora.
- Circuitos lógicos.
- Repertorio de instrucciones.
- Tipos de datos y formato de representación a nivel máquina.
- Ciclo de la instrucción.
- Programación en lenguaje ensamblador.
- Módulos de entrada y salidas y su comunicación con los periféricos.
- Gestión de interrupciones.
- Medidas de rendimiento del procesador.
- Jerarquía de memoria.
- Segmentación de instrucciones

Por otro lado, como parte de las actividades de difusión al medio del proyecto se realizó una charla sobre Simulación de Arquitecturas de Computadoras, en el mes de noviembre de 2021 en la escuela Técnica N° 2 Independencia de la Ciudad de Concordia a alumnos de 5º, 6º y 7º año. Además de contar con la presencia de varios docentes de la casa.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo del proyecto está conformado por 3 (tres) docentes de la carrera Licenciatura en Sistemas integrantes de la cátedra de Arquitectura de Computadoras y 1 (un) asesor externo. Adicionalmente este trabajo forma parte de 1 (una) tesis en curso de la Maestría en Sistemas de Información, dictada en la Facultad de Ciencias de la Administración, UNER. Un integrante está dirigiendo dos Tesis de Maestría en la Facultad de Ciencias de la Administración, Universidad Nacional de Entre Ríos. Finalmente, otro de los docentes investigadores de este proyecto continúa con su formación y perfeccionamiento.

5. BIBLIOGRAFÍA

- [1] J. Banks, J. S. Carson, B. L. Nelson, y D. M. Nicol, *Discrete-event system simulation*, 5th ed. Prentice Hall, 2010.
- [2] S. Robinson, *Simulation: The Practice of Model Development and Use*, 2nd edition. 2014.
- [3] A. M. Law, *Simulation Modeling & Analysis*, 5.a ed. New York, NY, USA: McGraw-Hill, 2015.
- [4] C. Lion, «Los simuladores. Su potencial para la enseñanza universitaria», *Cuad. Investig. Educ.*, vol. 2, n.o 12, pp. 53–66, 2005.
- [5] G. Contreras, R. G. Torres, y M. S. R. Montoya, «Uso de simuladores como recurso digital para la transferencia de conocimiento», *Apert. Rev. Innov. Educ.*, vol. 2, n.o 1, pp. 86–100, 2010.
- [6] B. Nova, J. C. Ferreira, y A. Araújo, «Tool to support computer architecture teaching and learning», en *Engineering Education (CISPÉE)*, 2013 1st International Conference of the Portuguese Society for, 2013, pp. 1–8.
- [7] B. Mustafa, «Evaluating A System Simulator For Computer Architecture Teaching And Learning Support», *Innov. Teach. Learn. Inf. Comput. Sci.*, vol. 9, n.o 1, pp. 100-104, 2010.
- [8] Z. Radivojevic, M. Cvetanovic, y J. Dordevic, «Design of the simulator for teaching computer architecture and organization», en *2011 Second Eastern European Regional Conference on the Engineering of Computer Based Systems*, 2011, pp. 124–130.
- [9] B. Nikolic, Z. Radivojevic, J. Djordjevic, y V. Milutinovic, «A Survey and Evaluation of Simulators Suitable for Teaching Courses in Computer Architecture and Organization», *IEEE Trans. Educ.*, vol. 52, n.o 4, pp. 449-458, nov. 2009.
- [10] R. Hasan y S. Mahmood, «Survey and evaluation of simulators suitable for teaching for computer architecture and organization Supporting undergraduate students at Sir Syed University of Engineering & Technology», en *Control (CONTROL)*, 2012 UKACC International Conference on, 2012, pp. 1043–1045.
- [11] J. L. Hennessy y D. A. Patterson, *Computer architecture: A quantitative approach*, Fifth Edition. Elsevier, 2012.
- [12] W. Stallings, *Computer organization and architecture: designing for performance*, Eleventh Edition. Pearson, 2013.
- [13] P. BEHROOZ, *Computer Architecture From Microprocessors to Supercomputers*. McGraw Hill, 2007.
- [14] Intel, «64 and IA-32 architectures software developers manual», 325462-060US, vol. 1, 2A, 2B, 2C, 2D, 3A, 3B, 3C and 3D, p. 4670, 2016.
- [15] AMD, «Developer Guides, Manuals & ISA Documents». [En línea]. Disponible en: <https://developer.amd.com/resources/developer-guides-manuals/>. [Accedido: 21-abr-2019].
- [16] P. Abel, *IBM PC Assembly Language and Programming*, Fifth Edition. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2000.
- [17] D. Skrien, «CPU Sim 3.1: A tool for simulating computer architectures for computer organization classes», *J. Educ. Resour. Comput. JERIC*, 2001.
- [18] A. Akram y L. Sawalha, «A comparison of x86 computer architecture simulators», 2016.

Evolución temática de publicaciones en español. Su aplicación en el campo de la tecnología y la educación

Santiago Bianco¹, Laura Lanzarini², Alejandra Zangara²

¹ Grupo de Investigación en Sistemas de Información UNLa (GISI-UNLa)

² Instituto de Investigación en Informática LIDI (UNLP-CIC)
sbianco@unla.edu.ar, {laural,azangara}@lidiinfo.unlp.edu.ar

CONTEXTO

Esta presentación corresponde a las tareas de investigación relacionadas con las tareas de análisis, búsqueda y recuperación de documentos que se llevan a cabo en el LIDI en el marco del proyecto “Sistemas inteligentes. Aplicaciones en reconocimiento de patrones, minería de datos y big data” perteneciente al Programa de Incentivos (2018-2022).

RESUMEN

La recuperación de información a partir de documentos de textos de un tema específico es un tema de sumo interés en distintas áreas. Un claro ejemplo puede observarse en la investigación científica. El inicio de una línea de investigación en un área específica por lo general comienza con una búsqueda del estado del arte de un tema de interés. A veces, incluso, es necesario identificar cuáles son los posibles temas de interés dentro de un dominio en particular. Este proceso de búsqueda bibliográfica generalmente consume mucho tiempo y si no se está bien guiado puede conducir a bloqueos y frustración para el investigador. Sería interesante entonces contar con métodos y herramientas que permitan simplificar la búsqueda y el análisis de bibliografía o publicaciones de cualquier tipo que facilite estos procesos.

Identificar los cambios en la terminología y la evolución en los temas de estudio resulta de sumo interés para disciplinas como la bibliometría y la cienciometría. De las mismas se destacan las técnicas de mapa temático y evolución temática, las cuales permiten extraer los temas de interés de una disciplina

en distintos períodos de tiempo, a través del análisis de publicaciones científicas o textos relevantes.

El problema principal de estas técnicas es que su implementación requiere conocimientos en minería de textos, procesamiento de datos, análisis estadístico, entre otras áreas específicas, lo que dificulta su uso por parte de usuarios no expertos.

Como los usuarios finales pueden ser investigadores, docentes, o cualquier persona que quiera hacer un relevamiento de una disciplina, en esta línea de investigación se propone acercar estas técnicas a usuarios no expertos a través de una metodología que pueda ser implementada sin conocimientos específicos en análisis de datos y afines.

Palabras clave: Análisis Bibliométrico, minería de Texto, Evolución Temática

1. INTRODUCCIÓN

El análisis de documentos de textos provenientes de contextos específicos es un tema de interés para distintas áreas tales como recuperación de información, clasificación de documentos, análisis bibliométricos y cienciométricos, entre otros.

Cuando se procesan documentos de textos escritos en períodos de tiempo diferentes, la evolución temática es un aspecto que debe tenerse en cuenta. Reconocer los cambios, que se han ido produciendo a lo largo del tiempo, en la denominación de los distintos tópicos dentro a una misma disciplina o área de discurso, es una herramienta sumamente útil a

la hora de querer aplicar estrategias pertenecientes a la Minería de Textos.

Como caso concreto, cualquier docente, investigador o estudiante que requiera hacer un artículo, tesis o trabajo de investigación, deberá efectuar una revisión del estado del arte correspondiente. En esta dirección, suele ser necesario identificar cuáles son los posibles temas de interés dentro de un dominio en particular. Este proceso de búsqueda bibliográfica generalmente consume mucho tiempo y si no se orienta correctamente puede conducir a bloqueos y frustración para el investigador. Sería interesante entonces contar con métodos y herramientas para simplificar la búsqueda y el análisis de bibliografía o publicaciones de cualquier tipo que faciliten estos procesos.

En primera instancia podrían usarse herramientas pertenecientes a la bibliometría para comenzar a analizar textos de interés. La Bibliometría se conoce como una disciplina capaz de describir un conjunto de publicaciones aplicando técnicas de análisis estadístico, identificando focos temáticos relevantes, redes de colaboración de autores, información sobre citas y demás. La Cienciometría es una subdisciplina de la Bibliometría que se enfoca particularmente en publicaciones científicas.

De todas formas, generalmente estos enfoques posibilitan análisis cuantitativos como por ejemplo una lista de autores más citados, instituciones que más publicaron, temas sobre los cuáles se escribió más, entre otros. En las ocasiones en que se requiera un análisis cualitativo más profundo deberán aplicarse técnicas de Minería de Textos y de visualización como mapas temáticos en conjunto con los métodos tradicionales de Bibliometría.

Los mapas temáticos son una forma de representar diferentes temas tratados en un campo de una disciplina científica en un determinado momento. Distinto tipo de información bibliométrica puede ser usada para armar estos gráficos, siendo una de ellas el análisis y la correlación entre términos relevantes.

De los mapas temáticos se deriva una técnica de análisis denominada evolución temática. La misma consiste en mostrar en una línea de tiempo la “evolución” de la relevancia de un tema en particular. Por ejemplo, se podría mostrar que en el 2010 hubo un foco temático dedicado a la investigación en Redes Neuronales y el mismo grupo de gente que trabajaba en ese tema fue inclinándose su investigación hacia otro distinto como puede ser la interpretabilidad de modelos de caja negra. La idea es mostrar que el primer tema mutó o evolucionó hacia el otro. Cabe destacar que en este contexto la palabra evolución denota cambio y transformación, y no necesariamente decir que “el Tema A evolucionó al Tema B” significa que el tema B es mejor en algún aspecto con respecto al Tema A.

Si bien existen algunas herramientas que permiten estos análisis, muy pocas se encuentran disponibles para el lenguaje español y no son tan intuitivas como para que puedan ser usadas por usuarios no expertos en el área informática. Además, están orientadas a publicaciones científicas en inglés descargadas de portales tales como Web of Science o Scopus, con un formato particular.

Por este motivo, se propone el desarrollo de una herramienta software que permita no solo automatizar el proceso sino facilitar la ejecución de la metodología descrita, por usuarios no expertos.

La herramienta en principio debe ser capaz de:

- Descargar los datos de revistas y formatearlos apropiadamente
- Cargar datos de revistas ya formateados para ser analizados por la herramienta
- Procesar y filtrar automáticamente datos faltantes u outliers
- Generar los gráficos correspondientes a la evolución temática de los textos.
- Permitir la visualización de los mapas temáticos generados para cada subperíodo analizado en la evolución temática.

- Filtrar el gráfico de la evolución temática para visualizar los resultados más relevantes

Además de los ya mencionados, un requisito fundamental de la herramienta es que pueda ser utilizada por usuarios no expertos. Esto implica automatizar lo más posible la aplicación de la metodología y reducir al mínimo la parametrización de los algoritmos utilizados. Idealmente sólo sería necesario que el usuario elija los años de corte para los subperíodos a analizar ya que el software se encargaría de elegir la mejor configuración entre los posibles parámetros a utilizar.

Esto en principio se utiliza una estrategia grid search sobre los mismos buscando maximizar el índice de inclusión en las relaciones entre términos encontradas, aunque se espera también experimentar con otros métodos como optimización bayesiana y algoritmos genéticos

Finalmente, la interfaz debe ser sencilla y directa, explicando los resultados obtenidos claramente sin aplicar fórmulas o detalles de implementación.

Debido a estos requisitos de usabilidad de la herramienta, la misma debe ser validada por usuarios para verificar que efectivamente no se necesitan conocimientos específicos para utilizarla. Además, se requiere que expertos validen los resultados finales que se obtengan a través de su uso. Es probable entonces que se requieran modificaciones luego de dichas validaciones, por lo que esto debe ser tenido en cuenta para la metodología de desarrollo a utilizar.

2. TEMAS DE INVESTIGACIÓN Y DESARROLLO

- Estudio de diferentes técnicas de representación de documentos. Se analizan representaciones de distinta complejidad que van desde el solo uso de las palabras clave provistas por los autores para luego incorporar el resto de los metadatos incluyendo las referencias bibliográficas.
- Estudio de distintas técnicas de Minería de Datos para determinar los temas de interés.

El énfasis está puesto en las técnicas de clustering aplicadas sobre distintas representaciones con el objetivo de comparar los temas identificados.

- Estudio e implementación de distintas técnicas de visualización como herramientas para analizar la relación de temas en un período específico y entre distintos períodos.
- Metodologías de desarrollo de software apropiadas para la herramienta.
- Mecanismos de relevamiento de las opiniones de los expertos en referencia a los resultados obtenidos.

3. RESULTADOS OBTENIDOS

- Diseño de técnicas apropiadas para analizar la evolución temática de publicaciones en español.
- Definición de métricas para filtrar los resultados más relevantes en un análisis de evolución temática.
- Desarrollo e implementación de una metodología para el análisis de la evolución temática de publicaciones en español.
- Diseño de una herramienta software capaz de implementar la metodología antes mencionada.
- Diseño e implementación de dispositivos de evaluación de la herramienta por parte de expertos.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo de la línea de I/D aquí presentada está formado por 2 profesoras doctoras con dedicación a la investigación y 1 tesista de maestría.

Dentro de los temas involucrados en esta línea de investigación, en los últimos 3 años se han finalizado 1 tesis de doctorado, 1 tesis de especialista y 1 tesina de grado de Licenciatura.

Actualmente se está desarrollando 1 tesis de maestría. También participan en el desarrollo de las tareas becarios y pasantes del III-LIDI.

5. REFERENCIAS

1. H. Alibrahim and S. A. Ludwig. Hyperparameter optimization: Comparing genetic algorithms against grid search and bayesian optimization. In 2021 IEEE Congress on Evolutionary Computation (CEC), pages 1551–1559. IEEE, 2021.
2. S. Bianco, L. C. Lanzarini, and M. A. Zangara. Evolución temática de publicaciones en español. una estrategia posible para el diseño de situaciones didácticas. In XVI Congreso de Tecnología en Educación y Educación en Tecnología (TEyET 2021). RedUNCI, 2021.
3. S. Bianco, L. C. Lanzarini, and M. A. Zangara. Thematic evolution of scientific publications in spanish. In XXVII Congreso Argentino de Ciencias de la Computación(CACIC)(Modalidad virtual, 4 al 8 de octubre de 2021), 2021.
4. M. Callon, J.-P. Courtial, and F. Laville. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, 22:155–205, 1991.
5. M. Callon, J.-P. Courtial, W. A. Turner, and S. Bauin. From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, 22(2):191–235, 1983.
6. M. Cobo, A. López-Herrera, E. Herrera-Viedma, and F. Herrera. An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, 5(1):146–166, 2011.
7. N. Coulter, I. Monarch, and S. Konda. Software engineering as seen through its research literature: A study in co-word analysis. *Journal of the American Society for Information Science*, 49(13):1206–1223, 1998.
8. P. K. Project. Open journal system, 2001. <http://pkp.sfu.ca/ojs>.
9. J. Ramos et al. Using tf-idf to determine word relevance in document queries. In Proceedings of the first instructional conference on machine learning, volume 242, pages 29–48. Citeseer, 2003.
10. S. Robertson. Understanding inverse document frequency: on theoretical arguments for idf. *Journal of documentation*, 2004.
11. C. Sternitzke and I. Bergmann. Similarity measures for document mapping: A comparative study on the level of an individual scientist. *Scientometrics*, 78:113 – 130, 2009.

Dispositivos móviles en colegios de nivel secundario

Pizarro Rubén, Testa Oscar, Camiletti Pablo, Ascheri María Eva, Ayala Ludmila
Departamento de Matemática / Facultad de Ciencias Exactas y Naturales/ Universidad
Nacional de La Pampa
Uruguay 151, 00-54-02954-245220 int 7124
rubenpizarro71@gmail.com ; otesta@gmail.com

RESUMEN

Los dispositivos móviles están modificando los ámbitos de enseñanza y aprendizaje, provocando importantes cambios en las formas de hacer y aprender de todos. Es necesario un rediseño metodológico y pedagógico para el abordaje de la educación de nivel secundario considerando el rol de estos nuevos dispositivos. Para estudiar este problema surge nuestro proyecto de investigación aprobado y financiado por la Facultad de Ciencias Exactas y Naturales de la UNLPam.

En el marco del proyecto se realizó una revisión bibliográfica sobre el tema de investigación, las aplicaciones existentes destinadas a la enseñanza en el nivel de la educación secundaria de las diferentes asignaturas y las experiencias de inclusión de dispositivos móviles en este nivel. Se obtuvo información para avanzar con la experiencia directa en el ámbito educativo. Además hemos desarrollado capacitaciones a docentes de nivel medio, vinculadas con la inclusión de dispositivos móviles en educación secundaria. Nos proponemos analizar la compleja influencia del acceso a las redes y la utilización de dispositivos móviles y cómo aprovechar estas características para experimentar, analizar y proponer mejoras en las formas de utilizar estas tecnologías en los colegios de nivel secundario de la Ciudad de Santa Rosa. Estos últimos años el desarrollo de las actividades se

vio afectada por la pandemia debiendo aún analizar y elaborar conclusiones.

Palabras clave: dispositivos móviles, aplicaciones, enseñanza aprendizaje.

CONTEXTO

El proyecto de investigación que se presenta está radicado en el Departamento de Matemática de la Facultad de Ciencias Exactas y Naturales de la UNLPam y financiado por dicha Facultad.

Tiene una duración de cuatro años, fue prorrogado en 2021 y 2022 a causa de la pandemia. Está vinculado con otros proyectos del Departamento ya ejecutados y en ejecución, relacionados con la enseñanza en nivel secundario y la inclusión de tecnologías en educación.

1. INTRODUCCIÓN

El artefacto simbólico omnipresente en la mayor parte de nuestras mediaciones hoy en día es el celular o dispositivo móvil, que evidentemente han modificado el entorno de estudiantes y profesores, provocando profundos cambios en las formas de hacer y aprender de todos. Precisamente porque observamos que los modos de conocer están íntimamente vinculados con las personas, los artefactos disponibles y las relaciones

establecidas entre personas a través del uso de artefactos, es que proponemos encontrarnos con docentes en las escuelas para indagar las formas en las que ellos conviven con estos aparatos y las redes de relaciones que con ellos construyen, los contenidos que comparten y producen. Estos dispositivos que surgieron como herramientas de comunicación y entretenimiento han llegado a desarrollar un papel fundamental en el mundo de la economía y el conjunto de la sociedad (Shuler et al, 2013).

Sin duda los dispositivos móviles están presentes en la vida de la mayoría de docentes y estudiantes para realizar diversas actividades. Ante esta realidad los Ministerios de Educación en diferentes países han iniciado políticas tendientes a que estos dispositivos cobren mayor importancia también en educación. En esta línea podemos mencionar los modelos 1 a 1 implementados en muchos países que consisten en la provisión por parte del estado de un dispositivo móvil (netbook o tableta) a cada alumno y docente. También existen iniciativas del tipo “Trae tu propia tecnología” denominada (BYOT, por sus siglas en inglés ‘Bring Your Own Technology’), consistentes en que cada estudiante lleve su propio dispositivo móvil para utilizarlo en la escuela. Esta metodología se ha implementado en sectores sociales más desarrollados en los cuales todos pueden acceder a estos dispositivos.

Los proyectos mencionados anteriormente, entre otros de similares características, permiten sin dudas un mayor acceso a la tecnología y aumentan las posibilidades de acceso a la información de los docentes y estudiantes. Permitirían así acceder a las principales características tecnológicas del

aprendizaje móvil que según indica Cantillo Valero, et. al. (2012), son las siguientes:

- Portabilidad, debido al pequeño tamaño de los dispositivos.
- Inmediatez y conectividad mediante redes inalámbricas.
- Ubicuidad, ya que se libera el aprendizaje de barreras espaciales o temporales.
- Adaptabilidad de servicios, aplicaciones e interfaces a las necesidades del usuario. También existe la posibilidad de incluir accesorios como teclados o lápices para facilitar su uso.

Vivimos indudablemente momentos en los cuales las tecnologías avanzan vertiginosamente pero existe una amplia brecha entre el desarrollo y la implementación (Montoya, 2009); es así que continuando con el análisis de este autor, nos indica que para lograr la inclusión de los dispositivos móviles, en especial los teléfonos inteligentes, se puede temporizar tres pasos:

1. Apropiación del objeto
2. Apropiación de la funcionalidad
3. Apropiación de la nueva forma de aprendizaje

Las dos primeras etapas como señala Díaz Bilbao (2014), serían alcanzadas rápidamente por los estudiantes, quedando la tercera etapa a cargo de las instituciones educativas y los docentes.

Sin duda la mayoría de los autores coinciden en indicar que la inclusión de dispositivos móviles en el aula, produce grandes beneficios: motivan a los estudiantes, aumenta las habilidades sociales mejorando la cooperación y la colaboración mejorando la creatividad y la capacidad cognitiva.

Sin embargo, como ha sucedido con otras tecnologías, percibimos que los dispositivos móviles actualmente no son tenidos en cuenta

ni aprovechados en las formas ni en la magnitud que observamos sería conveniente. Por el contrario en algunos ámbitos han sido prohibidos por causar distracción y otros problemas entre los estudiantes y la comunidad educativa.

Como lo expresan Castells, Palamidessi y muchos otros, los cambios que se han producido en las formas de comunicarse, acceder a información y conocer el mundo, mucho tienen que ver el acceso a las redes y los nuevos dispositivos y esto pareciera ser de alguna manera todavía resistido en el ámbito educativo. Al mismo tiempo se reconoce una crisis identificada en las dificultades que se presentan con la falta de atención, falta de aprendizaje o productividad, de motivación, de malestar de estudiantes y profesores entre otros problemas.

¿Tendremos que plantearnos abordar el trabajo en las instituciones educativas desde nuevos paradigmas? Pensar alternativas para las instituciones educativas implica salir de lugares comunes, observar que el mundo y la sociedad han cambiado mucho, en comparación las instituciones educativas, emblemáticas a la hora de hablar del conocimiento, parecen no haber dado signos claros de entender la metamorfosis del conocimiento a la que hoy asistimos.

A todo esto debemos sumar las experiencias que incorporó la pandemia en cuanto a las diferentes modalidades de desarrollo de las clases y las diferentes herramientas utilizadas entre las cuales los dispositivos móviles cobraron gran protagonismo. Sugiriéndonos así la implementación de estrategias heterogéneas para contextos heterogéneos como indica Marés L (2021)

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las líneas de investigación que se abordan con el presente proyecto son las siguientes:

- Nivel de penetración de los dispositivos móviles en las actividades de enseñanza aprendizaje en aulas de nivel medio.
- Búsqueda, experimentación y análisis de aplicaciones educativas para la enseñanza en el nivel medio
- Características metodológicas de las clases de nivel medio en las que se incluyen dispositivos móviles.

3. RESULTADOS OBTENIDOS / ESPERADOS

Dado que este proyecto nos lleva a insertarnos en la escuela secundaria de la ciudad de Santa Rosa provincia de La Pampa, no dudamos en hacer un relevamiento propio de los docentes de alguna de las instituciones en las que estamos en contacto. Este relevamiento lo hicimos mediante un cuestionario cuyas dimensiones de análisis fueron las características de los teléfonos celulares de los docentes, la frecuencia y tipo de utilización de su teléfono celular y los teléfonos celulares en la institución educativa.

Claramente la utilización de dispositivos móviles no es una actividad ajena a los docentes.

Hemos desarrollado acciones de capacitación para docentes de nivel medio sobre la utilización de estas tecnologías en las aulas. A partir de dichas capacitaciones se puede concluir que los docentes conocen un conjunto de aplicaciones con características generales que le permiten el manejo de archivos en el celular, el registro de sonido, capturas de

pantalla y registrar videos con el funcionamiento de aplicaciones. También pueden obtener y utilizar un conjunto de aplicaciones específicas de cada una de las áreas como pueden ser ciencias exactas, naturales o sociales. Sin embargo siguen existiendo dificultades en cuanto a la metodología a implementar para la utilización concreta del celular como una herramienta para el desarrollo de actividades y evaluaciones. También la diversidad de características de los equipos como las posibilidades de acceso a internet constituyen limitaciones al momento de concretar su utilización

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo es de carácter interdisciplinario, está conformado con especialistas del área de Educación, Matemática y Computación.

Se realizan, también, actividades de capacitación y transferencia referidas a la utilización de aplicaciones educativas para dispositivos móviles con sistema operativo Android.

5. BIBLIOGRAFÍA

- Castells, M. (2016). ¿Comunidades virtuales o sociedad en red?. Plaza & Janés Editores. Barcelona, España
- Castells, M., & Schmalenberger, M. (2006). *La sociedad red: una visión global* (No. U10 1087). IICA, Bs As (Argentina).
- Cantillo Valero, C., Roura Redondo, M. & Sánchez Palacín, A. (2012). Tendencias actuales en el uso de dispositivos móviles en educación. Educational Portal of the Americas – Department of Human Development, Education and Culture. OEA-OAS ISSN 0013-1059 La Educ@ción Digital Magazine N 147 – www.educoas.org
- Díaz Bilbao (2014) El smartphone como herramienta educativa. Trabajo final de maestría. Máster en formación del profesorado de educación secundaria (ámbito tecnológico). Universidad Pública de Navarra.
- Marés L. (2021). Escenarios combinados para enseñar y aprender: escuelas, hogares y pantallas / Educ.ar S.E.; 1a ed. Ciudad Autónoma de Buenos Aires. Disponible en <https://www.educ.ar/recursos/155488/>
- Marés L. (2021) Claves y caminos para enseñar en ambientes virtuales. Educ.ar S.E. - 1a ed. Ciudad Autónoma de Buenos Aires. Disponible en <https://www.educ.ar/recursos/155487/>
- Montoya, M. (2009). Mobile learning – mlearning- technology resources and their relationship with distance learning environments: applications and research studies. Tecnológico de Monterrey, ITESM (México).
- Miranda Ruiz, E. (2010). Paradigma Interpretativo en Investigación. Disponible <http://www.monografias.com/trabajos97/paradigma-interpretativo/paradigma-interpretativo.shtml>
- Palamidessi, M. (2006). La escuela en la sociedad de redes. Una introducción a las tecnologías de la información y la comunicación.

- RINALDI, M. (2012). Revolución Mobile Learning. Disponible en: <http://es.slideshare.net/crossmedialab/revolucion-mobile-learning> Consultado 06/08/2013.
- Shuler C, Niall Winters & Mark West. (2013) El futuro del aprendizaje móvil. Implicaciones para la formulación y planificación de políticas et al. UNESCO working paper series on mobile learning. Disponible en <http://unesdoc.unesco.org/images/0021/002196/219637s.pdf> Consultado 10/08/16.
- UNESCO (2012) Aprendizaje móvil para docentes. Temas globales. Recuperado el 17 de marzo de 2017 de <http://unesdoc.unesco.org/images/0021/002164/216452s.pdf>

HERRAMIENTAS DIGITALES PARA EDUCACIÓN. ANÁLISIS DE SU ACCESIBILIDAD

Javier Díaz, Alejandra Schiavoni, Paola Amadeo, Ivana Harari,
Soledad Gómez, Alejandra Osorio

LINTI - Laboratorio de Investigación en Nuevas Tecnologías Informáticas.

Facultad de Informática. Universidad Nacional de La Plata

Calle 50 esq. 120, 2do Piso. Tel: +54 221 4223528

jdiaz@unlp.edu.ar, ales@info.unlp.edu.ar, pamadeo@linti.unlp.edu.ar, iharari@info.unlp.edu.ar,
sgomez@cespi.unlp.edu.ar, aosorio@cespi.unlp.edu.ar

RESUMEN

El presente artículo describe el estudio de un conjunto de herramientas digitales utilizadas en educación, especialmente en el nivel superior, para analizarlas según sus características de accesibilidad. La pandemia y la situación de aislamiento virtualizaron en forma muy rápida las propuestas educativas presenciales. La modalidad a distancia nos enfrentó a la revisión de las prácticas docentes incluyendo la planificación, evaluación y seguimiento de los estudiantes y a una evaluación exhaustiva de las herramientas a utilizar. En este período, fue necesario el uso de plataformas educativas y de recursos accesibles para garantizar que el estudiante pueda interactuar con el docente y el material de estudio. Además, las herramientas de videoconferencia emergieron como el único recurso que permitió los encuentros sincrónicos y la interacción en forma colaborativa. El objetivo es analizar las funcionalidades provistas para discapacidad, su cumplimiento y conformidad de estándares internacionales de accesibilidad. Se hace mención a diferentes herramientas que permiten validar las plataformas utilizadas y crear material accesible para los estudiantes.

Palabras clave: accesibilidad, herramientas educativas, recursos accesibles.

CONTEXTO

En el LINTI, Laboratorio de Investigación en Nuevas Tecnologías Informáticas de la Facultad de Informática de la UNLP, se vienen llevando a cabo varias líneas de investigación sobre Accesibilidad Web y su aplicación en diferentes ámbitos, considerando plataformas de aprendizaje abiertas. En la Facultad de Informática se utilizan plataformas de código abierto para las gestiones académicas desde hace más de quince años, incluyendo sistemas de gestión de aprendizaje como Moodle, repositorios abiertos y sistemas de gestión administrativa, como SIU Guaraní. También se realizan continuos mecanismos de integración entre las diferentes plataformas, lo que permite una mayor flexibilidad y aprovechamiento en el uso de las mismas.

Las líneas de investigación que se mencionan en este artículo se desarrollan en LINTI de la UNLP y está enmarcado en el proyecto I+D 11/F028 “De la Sociedad del Conocimiento a la Sociedad 5.0: un abordaje tecnológico y ético en nuestra región”, aprobado en el marco del

Programa de Incentivos, bajo la dirección del Lic. Javier Díaz. Este proyecto hace hincapié entre otros puntos en la formación de competencias y habilidades digitales para todos los ciudadanos, en una sociedad que se plantea como digital.

Dada su relevancia, el tema de accesibilidad web se viene trabajando en la Facultad desde el año 2002, y se incorporó esta temática en el plan de estudios de las carreras que se dictan en la institución, a través de la asignatura Diseño Centrado en el Usuario. El libro “Guía de recomendaciones para diseño de software centrado en el usuario” registra aproximadamente 10.000 descargas desde el año 2013 [1]. También se institucionalizó su abordaje mediante la creación de una Dirección de Accesibilidad desde el año 2010, se desarrollaron tesinas, trabajos de cátedra al respecto, proyectos de innovación y desarrollo con estudiantes de Informática y diferentes proyectos de extensión acreditados por la Universidad Nacional de La Plata. Los proyectos de los últimos años son “Trabajando por una Web Accesible”, “Por una Web Inclusiva”, “Por una Web inclusiva: abordaje en escuelas secundarias”, que entró en el Programa de promoción de derechos y fortalecimiento de la organización comunitaria [2]. También, se aprobó el Proyecto de Desarrollo e Innovación sobre “Rampas Digitales Innovativas para Personas con Discapacidad”, en la convocatoria Agregando Valor 2018-2019, del Ministerio de Ciencia, Tecnología e Innovación de la Nación Argentina.

1. INTRODUCCIÓN

La situación de aislamiento provocada por la pandemia, hizo que fuera no sólo indispensable sino urgente adecuar las

estrategias de enseñanza en todos los niveles educativos. La modalidad totalmente a distancia, que experimentamos por primera vez, nos enfrentó a la revisión de nuestras prácticas docentes, incluyendo la planificación general de la cursada, evaluación, seguimiento y otros aspectos que involucran el proceso de enseñanza-aprendizaje.

Al mismo tiempo, herramientas como la conferencia web, una herramienta bidireccional y sincrónica, emergió como el único recurso que nos permitió tener un encuentro simultáneo e interactuar en forma colaborativa. La envergadura, que estas herramientas cobraron durante los periodos de aislamiento, estuvo en relación a: un aspecto pedagógico, a la consolidación de vínculos, una potencialidad educativa, la comunidad virtual y un aspecto social, en tanto permitió la inclusión de participantes que se encuentran dispersos geográficamente. Hace algunos años los sistemas de conferencias web eran bastante limitados, pero actualmente se puede contar con sistemas interesantes de videoconferencias que ofrecen una diversidad de funcionalidades, constituyendo una herramienta con un potencial pedagógico significativo.

Si bien las herramientas de videoconferencia, en muchos casos potentes, nos permiten un determinado acercamiento, éste no es comparable con el presencial. El medio virtual, el software intermediario, los problemas de comunicación, la interferencia, los ruidos, hacen que dicha virtualidad se haga presente y la distancia se sienta. En el caso de las personas con discapacidad, las limitaciones que presentan las herramientas de videoconferencia (problemas de comunicación, la interferencia, los ruidos) se profundizan aún más, dificultando el desarrollo de

prácticas educativas que promuevan procesos de formación y aprendizaje. El no poder conectarse, o conectarse pero no poder acceder al material completo, o no poder escuchar lo que el docente está explicando, o no poder ver las diapositivas que el docente está compartiendo a través de la videoconferencia, agudiza aún más las distancias. Esta brecha genera un distanciamiento difícil de sortear si no es a través de la implementación de estrategias específicas, que promuevan alternativas consensuadas con los 1500 estudiantes con discapacidad. La inclusión de estas dinámicas, se configuran como puentes que la cátedra debe posibilitar, que deben ser construidos en conjunto con la persona con discapacidad atendiendo sus necesidades de comunicación, adaptando el material que provee y su forma de abordarlo, según la especificidad de su caso y de sus limitaciones que pueden ser mitigadas si las condiciones requeridas son otorgadas.

Desde esta perspectiva, es necesario contemplar las plataformas virtuales de aprendizaje y enseñanza, los recursos digitales, las páginas web de referencia, las TIC, los sistemas de videoconferencia que son utilizadas por las personas con discapacidad, entre otras cuestiones. En el presente artículo se reconoce la complejidad de la temática, analizando distintos productos, que abarcan los puntos mencionados previamente. Además de la plataforma virtual Moodle, se analizan y comparan los sistemas para videoconferencia: Zoom, Big Blue Button y Webex. Para ello, se realizaron una serie de testeos y comprobaciones, como, por ejemplo, las funcionalidades provistas para discapacidad, su cumplimiento y conformidad de estándares internacionales de accesibilidad. Todas las herramientas fueron testeadas

manualmente en distintos escenarios de interacción. Además, se presentan diferentes herramientas que permiten validar y crear material accesible para los estudiantes y permiten tener en cuenta un conjunto de consideraciones importantes a tener en cuenta sobre la accesibilidad principalmente desde el punto de vista de estos sistemas de videoconferencia que se están utilizando para suplantar las clases presenciales [3].

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

Dentro de la línea de trabajo que venimos describiendo, las tareas de investigación, desarrollo e innovación que se llevan a cabo en este proyecto incluyen los temas relacionados a la accesibilidad en las herramientas utilizadas en educación. El tema de accesibilidad web se trata en forma transversal, evaluando el cumplimiento de las normas respectivas en cada uno de los casos.

Como se mencionó anteriormente, debido al contexto de salud, dio lugar a restricciones de movilidad y de confinamiento social preventivo, que se tuvo que implementar de manera obligatoria y rigurosa en la mayoría de los países, y Argentina no fue la excepción. El uso de dispositivos y medios tecnológicos digitales, específicamente de la Web, se convirtió en un medio demandante y obligatorio, principalmente para continuar en el proceso de enseñanza y aprendizaje.

Las clases presenciales se convirtieron en encuentros sincrónicos a través de sistemas de videoconferencias, las consultas en foros, las instancias evaluativas en sistemas de examen remoto, con cámaras y micrófonos encendidos, intentando que el proceso

educativo no se interrumpa y pueda establecerse.

En el caso de las personas en situación de discapacidad, no sólo basta con que las clases, los materiales y demás recursos educativos estén disponibles en Internet, sino que además dichos medios y herramientas digitales provistas, sean accesibles.

Por esto, en este proyecto el estudio e investigación apunta a analizar el nivel de accesibilidad de las herramientas digitales que se utilizan para realizar las clases sincrónicas remotas, como lo son los sistemas de videoconferencias, más allá del cumplimiento de estándares internacionales de accesibilidad mencionados por las propias herramientas desde la perspectiva de estudiantes de educación superior con distintas discapacidades. También, Diseñar un curso accesible requiere pensar en un curso que sea accesible para cualquier tipo de estudiante, es decir, sean capaces de recibir información, interactuar con el material de estudio, con el docente y sus pares y demostrar su aprendizaje más allá de su edad, género, raza, lengua, discapacidad, etc. El framework de diseño de aprendizaje universal [4] es una guía de recomendaciones para generar un entorno de aprendizaje flexible que apunta a mejorar la percepción, la expresión y la comprensión. En este sentido, las plataformas digitales de aprendizaje-enseñanza, el material de estudio, los canales de comunicación y todas las herramientas deben ser accesibles. Brindar múltiples opciones no implica bajar las expectativas, sino generar un espacio inclusivo garantizando que el aprendizaje se lleve a cabo para todos los participantes.

3. RESULTADOS Y OBJETIVOS

Debido al confinamiento impuesto en este tiempo, el uso de dispositivos y medios tecnológicos digitales, específicamente de la Web, se convirtió en un medio demandante y obligatorio, principalmente para continuar en el proceso de enseñanza y aprendizaje.

Según las líneas de trabajo descritas que se vienen desarrollando hace varios años, se plantean los siguientes objetivos:

- Análisis de sistemas de videoconferencia:
 - Utilizando herramientas automáticas como Wave [5]
 - A través de validaciones manuales con lectores de pantalla como NVDA [6]
 - Con herramientas específicas para chequear el contraste de colores, posibilidad de aumentar el tamaño de las letras [7] [8] [9]
- Validación de documentos, considerando distintas clases de documentos, realizados con editores de texto, presentaciones, documentos PDF:
 - Análisis de herramientas para creación y validación, tal como el Acobat Reader Pro [10], Pave [11], TingTungChecker [12], AccessibilityChecker [13]
 - Análisis de las normas ISO 14289-1:2014 PDF/UA [14] que describen los componentes requeridos y prohibidos, y las condiciones para su inclusión en un archivo PDF
- Análisis de pautas para la creación de videos accesibles
 - Estudio de las normas WCAG 2.1 incluye una serie de recomendaciones para contenido audiovisual [15]
 - Análisis de las herramientas para la grabación de las clases, tales

como OBS Studio [16], así como también complementos de navegadores, como la extensión Awesomescreenshot [17] o Loom [18].

- Análisis de los distintos formatos de subtítulos: Subrip, SMIL y WebVTT. Estudio de herramientas, tales como Express Scribe [19], herramienta de subtítulo para Windows y Mac, en español; CapScribe (en etapa de testeo) [20], Gnomesubtitles, herramienta de subtítulo gratuita para Linux [21]; Gaupol, una herramienta gratuita para Windows que permite subtítular [22], entre otras.
- Análisis de las normas de accesibilidad para la creación de contenido web, teniendo en cuenta las pautas a cumplir en el uso de tablas, links, imágenes y gráficos. según la WCAG 2.1 [23].

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo se encuentra formado por tres profesoras de amplia trayectoria en el campo de la investigación, que trabajan en el área de ambientes virtuales de aprendizaje y accesibilidad web. Además, una alumna becaria de la carrera de Licenciatura en Informática de la Facultad de Informática.

La participación en eventos de la especialidad, ha permitido seguir estableciendo canales de comunicación con otros investigadores que trabajan en las mismas áreas.

Las Jornadas de Accesibilidad que se realizan todos los años permiten intercambiar experiencias entre los participantes del curso y exponer los trabajos realizados. El video de la Jornada

del año 2021 se encuentra disponible en: <https://www.youtube.com/watch?v=3iUM2Yrg2TA>.

5. REFERENCIAS

- [1] Díaz, J., Amadeo, P., Harari, I. “Guía de recomendaciones para diseño de software centrado en el usuario”. Editorial de la Universidad Nacional de La Plata (EDULP). ISBN: 978-950-34-1030-1. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/32172>
- [2] Dictamen Proyectos de Extensión Acreditados en la UNLP. https://unlp.edu.ar/proyctosext/proyectos_de_extension_acreditados_y_subsidios-4708
- [3] Díaz J., Harari I., Schiavoni A., Amadeo A., Gómez S., Osorio, A.: “Aporte para pensar la educación en pandemia desde la accesibilidad.” Proceedings del Congreso Argentino en Ciencias de la Computación CACIC 2021 compilación de Marcia I. Mac Gaul. - 1a ed ISBN 978 -987-633-574-4 (2021).
- [4] CenterforTeachingInnovation. CornellUniversity. <https://teaching.cornell.edu/teaching-resources/designing-your-course/universal-design-learning>
- [5] WAVE Web AccessibilityEvaluationTool: <https://wave.webaim.org/>
- [6] NVDA en Español: <https://nvda.es/>
- [7] Web Developer. <https://chrome.google.com/webstore/detail/web-developer/bfbameneiokkgbdmiekhjnmfkcnldhbm>
- [8] ColourContrastChecker.

<https://chrome.google.com/webstore/detail/colour-contrast-checker/nmmjeclfkjdomacpcflgdkgpphpnmfe/related?hl=en-GB>

[9] AccessibilityInsightsfor Web. <https://accessibilityinsights.io/docs/en/web/overview/>

[10] Crear y verificar la accesibilidad de archivos PDF (Acrobat Pro) <https://helpx.adobe.com/es/acrobat/using/create-verify-pdf-accessibility.html>

[11] PAVE <https://pave-pdf.org/>

[12] TingtunChecker <https://checkers.eiii.eu/en/pdfcheck/>

[13] ImproveaccessibilitywiththeAccessibilityChecker<https://support.microsoft.com/en-us/office/improve-accessibility-with-the-accessibility-checker-a16f6de0-2f39-4a2b-8bd8-5ad801426c7f>

[14] ISO 14289-1:2014 DocumentmanagementapplicationsElectronicdocument file formatenhancementforaccessibility<https://www.iso.org/standard/64599.html>

[15] Howto Meet WCAG (Quick Reference) <https://www.w3.org/WAI/WCAG21/quickref/#time-based-media>

[16] OBS Studio <https://obsproject.com/>

[17] Awesomescreenshot <https://www.awesomescreenshot.com/>

[18] Loom <https://chrome.google.com/webstore/detail/loom-%E2%80%93-free-screen-and-cast-to-your-computer/liecbddmkiihnedobmlmillhodjkdmh?hl=en>

[19] Express Scribe Transcription Software.

<https://www.nch.com.au/scribe/index.html>

[20] CapScribe 2. <https://capscribe.ca/2.0/>

[21] GnomeSubtitles. <https://gnomesubtitles.org/>

[22] Gaupol. <https://otsaloma.io/gaupol/>

[23] Web Content AccessibilityGuidelines (WCAG) 2.1. <https://www.w3.org/TR/WCAG21/>

Objetos de Aprendizaje: diseño y construcción basado en criterios de calidad

Fernanda CARMONA¹, Antonio CASTRO LECHTALER^{1,2}, José TEXIER, Marisa GAGLIARDI¹,
Alberto Eduardo RIBA¹, Emmanuel PORTUGAL¹, Patricia MANRIQUES¹, José SORIA¹

¹Departamento de Ciencias Básicas y Tecnológicas, Universidad Nacional de Chilecito
9 de Julio 22, Chilecito, La Rioja, Argentina; ²CISTIC/FCE - Universidad de Buenos Aires, Córdoba 2122, CABA

{fbcarmona, jtexier, mgagliardi, ariba, eportugal}@undec.edu.ar, antonio.castro@fce.uba.ar,
patriciamanriques311995@gmail.com, soriaajose@gmail.com

RESUMEN

La incorporación de las Tecnologías de la Información y Comunicación (TIC) en la Educación ha permitido extender los ambientes de enseñanza y aprendizaje, para así poder desarrollar ambientes virtuales colaborativos e interactivos, con el uso de diferentes materiales educativos digitales que permitan el logro de los objetivos.

Adicionalmente, En los últimos años, el desarrollo de Repositorios Institucionales (RI) de acceso abierto ha sido un tema prioritario en las políticas de educación, ciencia y técnica de muchos países, y en particular, en las universidades públicas de Argentina, aprobándose la Ley 26899 de Repositorios Institucionales (RI), y en los otros niveles educativos las Bibliotecas Escolares con la aprobación de la Ley 26.917 sobre el Sistema Nacional de Bibliotecas Escolares y Unidades de Información Educativa.

Esta línea de I+D+i corresponde al diseño, desarrollo e implementación de proyectos que fortalezcan la investigación y las diferentes labores relacionadas con la gestión del conocimiento en cuanto a la visibilidad web y la preservación de la producción educativa y académica de la UNdeC y de las instituciones intervinientes.

Dentro de esta línea de trabajo se encuentran el desarrollo de Repositorios Institucionales (RI), Objetos de Aprendizaje (OA) y el movimiento de Acceso Abierto (AA); tanto en lo que respecta a los aspectos técnicos vinculados a ellos, como a su

utilización y uso como productos útiles en el proceso de enseñanza – aprendizaje en los diferentes niveles educativos.

Palabras clave: repositorios institucionales, objetos de aprendizaje, material educativo digital, acceso abierto.

CONTEXTO

Actualmente la UNdeC se encuentra trabajando sobre una política de fortalecimiento de las carreras en su modalidad presencial con el fin de poder llegar a lugares más distantes de la región, comenzando paso a paso a transitar el camino de la virtualidad con el desarrollo de nuevas carreras. A partir del año 2013 se implementó el campus virtual institucional en <http://campus.undec.edu.ar>, sobre la plataforma E-ducative, con aulas virtuales como soporte a la enseñanza presencial. Se crea el área de Tecnologías de la Información y la Comunicación en la Educación, que tiene a cargo la implementación de los servicios de educación a distancia, campus virtual UNdeC, sistemas de videoconferencias y capacitación y soporte para la inclusión de TIC en propuestas pedagógicas. Modalidad que se fue fortaleciendo durante todo el año 2020 y que posibilitó, durante del aislamiento social y preventivo impuesta por la pandemia por COVID-19, que los estudiantes pudieran mantener el cursado con regularidad, inclusive posibilitando que muchos de

ellos pudieran egresar con su título universitario.

En el año 2018 se inician las gestiones y el trabajo para la implementación del Repositorio Institucional adquiriendo durante el primer semestre 2019, un servidor de gran porte para tal fin.

En este marco esta línea de I+D+i corresponde al diseño, desarrollo e implementación de proyectos que fortalezcan la investigación y las diferentes labores relacionadas con la gestión del conocimiento en cuanto a la visibilidad web y la preservación de la producción educativa y académica de la UNdeC y de las instituciones intervinientes.

La Resolución Rectoral UNdeC N° 155/20 aprueba el financiamiento y ejecución del proyecto “Metodología para el diseño y construcción de objetos de aprendizaje basada en criterios de calidad” presentado en la convocatoria FICyT UNdeC (Financiamiento para estímulo y desarrollo de la Investigación Científica y Tecnológica), período de ejecución 2020-2022. El proyecto propone una metodología para el diseño y construcción de OA, que incluya técnicas de evaluación de calidad orientado a realizar el proceso de evaluación desde diferentes enfoques, definiendo un conjunto de características que debe cumplir el objeto, asociadas a diferentes visiones y actores que intervienen en el proceso de análisis, construcción y evaluación. Esta temática forma parte del proyecto de tesis de maestría de uno de los integrantes del equipo de trabajo que vienen trabajando en la UNdeC en colaboración con otras instituciones del país y del extranjero a través de los proyectos “Red para la Integración de Universidades en el uso de TIC para la Inclusión en la Educación Superior” aprobado en la 7 Convocatoria a Redes Internacionales, año 2013, el proyecto “Red para la creación y publicación de objetos virtuales de aprendizajes de calidad en Repositorios Institucionales” aprobado en la 9 Convocatoria a

Redes Internacionales, 2016-2017, y el Proyecto “Repositorios Digitales con Contenidos Orientados a las Necesidades de Escuelas Rurales (ER)” por Resolución CE N° 1055/15 convocatoria de Proyectos de Desarrollo Tecnológico y Social (PDTS), 2016-2018, presentado en el marco del Plan de Fortalecimiento de la Investigación Científica, el Desarrollo Tecnológico y la Innovación en las Universidades Nacionales (Ac. Pl. N° 676/08 y N° 687/09) y del Reglamento PDTS-CIN (Ac. Pl. N° 901/14), del Consejo Interuniversitario Nacional y el CONICET.

En el marco de esta línea de investigación en diciembre de 2020 se defendió y aprobó un Trabajo Final para la obtención del título de grado Licenciado/a en Sistemas denominado “Sistema de evaluación de métricas de calidad de Objetos de Aprendizaje”.

INTRODUCCIÓN

La incorporación de las Tecnologías de la Información y Comunicación (TIC) en la Educación ha permitido extender los ambientes de enseñanza y aprendizaje, para así poder desarrollar ambientes virtuales colaborativos e interactivos, con el uso de diferentes materiales educativos digitales que permitan el logro de los objetivos. Es por ello, que en el ámbito educativo se ha impuesto un nuevo concepto que busca la reutilización, permanencia, interoperabilidad, accesibilidad y compatibilidad de recursos digitales para el desarrollo de cursos y programas de formación en línea a través de la Web, los llamados Objetos de Aprendizaje (OA). Lo que distingue a un OA de un material educativo digital es la introducción de información autodescriptiva, expresada a través de los metadatos, conjunto de atributos o elementos que permiten describir al objeto, es decir son entidades generalmente entregadas a través de Internet y diseñadas con el fin de que sean utilizadas y reutilizadas en múltiples contextos educativos [1] [2].

En los últimos años, el desarrollo de Repositorios Institucionales (RI) de acceso abierto ha sido un tema prioritario en las políticas de educación, ciencia y técnica de muchos países, y en particular, en las universidades públicas de Argentina, aprobándose la Ley 26899 de Repositorios Institucionales (RI), y en los otros niveles educativos las Bibliotecas Escolares con la aprobación de la Ley 26.917 sobre el Sistema Nacional de Bibliotecas Escolares y Unidades de Información Educativa.

Los RI tienen como propósito recopilar, catalogar, gestionar, acceder, difundir y preservar información, permitiendo entre otras cosas el acceso libre y gratuito a todos los recursos que los conforman [3] [4].

El desarrollo de Repositorios Institucionales (RI) de acceso abierto ha sido un tema prioritario en las políticas de educación, ciencia y técnica de muchos países [5]. Un RI puede integrar a los OA en los denominados Repositorios de Objetos de Aprendizaje (ROA), bibliotecas digitales especializadas, orientados a facilitar la búsqueda y recuperación de los OA de manera que puedan ser utilizados en diversos ambientes educativos [6].

Sin embargo, no es suficiente con que estos materiales educativos estén disponibles, se requiere que cumplan con un alto nivel de calidad de modo que permita mayores posibilidades de lograr los objetivos de aprendizaje esperados en los estudiantes [7]. El factor más importante para alcanzar el éxito en la educación es el grado con el cual los docentes son capaces de producir actividades estructuradas, con una buena aplicación de la tecnología. Aquí es donde entran a jugar un papel importante los contenidos y su diseño en forma de OA.

Para optimizar la creación de los OA se han diseñado varias metodologías que apoyan su proceso de desarrollo y construcción [8]. Si bien se han propuesto enfoques que permiten evaluarlos teniendo en cuenta diferentes criterios y técnicas, no se cuenta con

modelos de evaluación generalizados que permitan identificar el nivel de calidad de estos recursos integrando diferentes dimensiones [7] [9].

La mayoría de las estrategias de evaluación propuestas se enfocan en pocos criterios y están orientadas a un conjunto específico de OA, son pocos los modelos de evaluación que integren evaluaciones automáticas con respuestas de usuarios y expertos, que permitan evaluar aspectos visuales, de contenido y pedagógicos, entre otros [9]. Estos involucran diferentes usuarios relacionados con los OA almacenados en repositorios y los resultados sirven para ordenar los resultados de búsqueda, es decir que se perfilan como herramientas de gestión de la calidad de OA publicados en repositorios.

En esta línea de I+D+i se propone una metodología para la construcción de OA, que está orientada a realizar el proceso de evaluación de calidad desde diferentes enfoques y durante el desarrollo de cada una de sus fases. Es decir, una metodología que se convierta en referente a la hora de planear y desarrollar OA aportando al proceso una organización interna, la identificación y selección de las competencias a desarrollar, así como el tipo de actividades cognitivas y de evaluación, garantizando la calidad de los OA construidos mediante la implementación de técnicas de evaluación de calidad tanto en el diseño instruccional y como en el diseño tecnológico de los OA en las diferentes fases de su desarrollo.

LÍNEAS DE INVESTIGACIÓN y DESARROLLO

- Objetos de aprendizaje.
- Repositorios institucionales y bibliotecas digitales.
- Gestión de la información y el conocimiento.
- Sistemas de información web y bases de datos.
- Índices bibliométricos.
- Interoperabilidad.

- Preservación digital.
- Recuperación de la información.

OBJETIVOS

El desarrollo de esta línea de investigación está permitiendo cumplir con los siguientes objetivos:

- Comprender los problemas y las tendencias asociadas a la producción abierta del conocimiento, el uso y la difusión.
- Fortalecer la visibilidad web de la UNdeC y de las instituciones intervinientes.
- Fomentar la formación de habilidades y conocimientos relacionados con los OA y los RI.
- Caracterizar las diferentes estrategias y técnicas de evaluación de calidad de los OA.
- Definir las fases, criterios de calidad y técnicas de evaluación de calidad de para la construcción de OA.
- Desarrollar una metodología para el diseño y construcción de OA basada en criterios de calidad.
- Entender y valorar el movimiento mundial de Acceso Abierto y señalar las vías de su materialización.
- Implementar un Repositorio Institucional para visualizar los diferentes recursos educativos y académicos que se produzcan.
- Depositar y preservar los producción educativa y académica de la UNdeC y de las instituciones intervinientes.
- Definir e implementar políticas de recopilación, distribución y mantenimiento para el funcionamiento del RI.

RESULTADOS

OBTENIDOS/ESPERADOS

- Curso a nivel de postgrado llamado “Gestión de la Información Académica y Científica”.
- Desarrollo de un proyecto de asignaturas sobre las Bibliotecas Digitales para

presentarse en dos universidades [10] y otras que lo puedan requerir.

- Desarrollo de talleres para la formación de habilidades y conocimientos relacionados con los RI, los OA y su construcción.
- Caracterización de las diferentes estrategias y técnicas de evaluación de calidad de los OA.
- Desarrollo de una metodología de construcción de OA que incluya técnicas de evaluación de calidad.
- Implementación de un repositorio de para la gestión los recursos educativos y académicos de la UNdeC y de las instituciones intervinientes.

FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está formado por docentes investigadores categorizados y otros en formación especializados bibliotecas digitales, repositorios institucionales y en enseñanza en entornos virtuales de aprendizaje y estudiantes de las carreras Ingeniería en Sistemas y Licenciatura en Sistemas.

Tres de los miembros del equipo poseen formación de postgrado: un Doctor en Ciencias Informáticas, un Magister en Informática y un Especialista en Entornos Virtuales de aprendizaje. Dos integrantes se encuentran en la etapa de desarrollo de su trabajo de tesis para obtener el título de Magister en Informática uno y otro de Magister en Enseñanza en Entornos Digitales. Tres miembros están realizando el trabajo final de la Especialización en inteligencia de datos orientada a Big Data.

En diciembre de 2020 una de las integrantes del equipo defiende y aprueba su Trabajo Final para la obtención del título de grado Licenciado/a en Sistemas, denominado “Sistema de evaluación de métricas de calidad de Objetos de Aprendizaje”, con la dirección y asesoramientos de los integrantes de este equipo de trabajo. Otro de

los integrantes es estudiante avanzado de la carrera Licenciatura en Sistemas.

BIBLIOGRAFÍA

- [1]. D. Wiley, "Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy," presented at the Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos, SPEDECE, 2001.
- [2]. E. Morales, F. García, A. Barron, A. Berlanga, and C. López, "Propuesta de Evaluación de Objetos de Aprendizaje," presented at the Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables (SPDECE), Barcelona, 2005.
- [3]. J. Texier, "Los repositorios institucionales y las bibliotecas digitales: una somera revisión bibliográfica y su relación en la educación superior," presented at the 11th Latin American and Caribbean Conference for Engineering and Technology - 2013, Cancun, México, 2013, p. 9.
- [4]. M. De Giusti, N. Oviedo, A. Lira, A. Sobrado, J. Martínez, and A. Pinto, "SEDICI – Desafíos y experiencias en la vida de un repositorio digital," *RENATA*, vol. 1, no. 2, pp. 16–33, Aug. 2011.
- [5]. Texier, J. (2016). Los repositorios institucionales y su importancia en las Universidades Venezolanas. *Revista Scitus*. Venezuela.
- [6]. R. McGreal, "A Typology of Learning Object Repositories", In *Handbook on Information Technologies for Education and Training* (pp.5–28).doi:10.1007/978-3-540-74155-8_1. 2008
- [7]. V. Morales Tabares, "Modelo por Capas para Evaluación de la Calidad de Objetos de Aprendizaje en Repositorios de Objetos de Aprendizaje", Instituto de Sistemas y Ciencias de la Decisión. 2013
- [8]. C. Sanz, L. Moralejo, & F. Barranquero, Curso de Doctorado "Metodología CROA", Universidad Nacional de la Plata, Argentina, 2014
- [9]. V. Morales Tabares, N. Duque Méndez & D. Ovalle Carranza, "Modelo por capas para evaluación de la calidad de Objetos de Aprendizaje en repositorios". *Revista electrónica de investigación educativa*, 19(3), 33-48. 2017.
- [10]. J. Texier, J. Zambrano, & F. Carmona, "Las Bibliotecas Digitales en el Currículum de las carreras de Ciencias de la Computación: una propuesta posible para Argentina y Venezuela", XII Congreso Argentino de Ciencias de la Computación CACIC 2016.

Explorando con Realidad Virtual Interactiva

Claudia Banchoff, Laura Fava, Sofía Martín, Facundo Díaz Gira, Agustín Aguirre

LINTI - Laboratorio de Investigación en Nuevas Tecnologías Informáticas.
Facultad de Informática. Universidad Nacional de La Plata
Calle 50 esq. 120, 2^{do} Piso. Tel: +54 221 4223528
{cbanchoff, lfava, smartin}@info.unlp.edu.ar, {facundodiazgira,
agus.aguirre10}@gmail.com

RESUMEN

Continuando una línea de investigación y desarrollo que se viene trabajando en el LINTI, Laboratorio de Investigación en Nuevas Tecnologías Informáticas de la UNLP (Argentina), en este artículo se presenta la concreción de los proyectos que utilizan Realidad Aumentada (RA) con fines sociales y el estado de avance de aquellos que utilizan Realidad Virtual (RV) para potenciar aplicaciones móviles. Dentro de esta línea se abordan aspectos relacionados al desarrollo de aplicaciones educativas interactivas y materiales didácticos, así como también una nueva línea relacionada con RV aplicada a la arquitectura.

En este artículo, se continúa con las líneas de I+D presentadas en WICC 2021[1], donde se analizaron y evaluaron entornos de desarrollo para aplicaciones de gamificación usando RA y RV y se presentan los avances en los desarrollos de aplicaciones usando RA y RV.

Palabras clave: juegos serios, proyectos sociales, realidad aumentada, realidad virtual, gamificación.

CONTEXTO

Esta línea de investigación incluye el desarrollo de aplicaciones lúdicas interactivas vinculadas a distintas problemáticas sociales. Se utilizan

tecnologías de RA y/o RV como elementos motivadores para complementar las actividades que se trabajan en distintos contextos.

En el LINTI se viene trabajando, desde hace más de 10 años, en proyectos relacionados a demandas de la sociedad. En el último año se finalizaron dos desarrollos que utilizan RA y se avanzó con un desarrollo de RV descrito más adelante en este artículo.

La línea de investigación que se presenta en este trabajo viene desarrollándose desde hace tres años y se encuentra enmarcada en el proyecto: "De la Sociedad del Conocimiento a la Sociedad 5.0: un abordaje tecnológico y ético en nuestra región", del Programa Nacional de Incentivos a docentes-investigadores, que se desarrolla en el LINTI.

1. INTRODUCCIÓN

La realidad aumentada (RA) y la realidad virtual (RV) se han utilizado en múltiples áreas, aunque en nuestra región, no ha sido sencilla su incorporación.

En el caso de RV, la adopción y aplicación es más compleja debido al requerimiento adicional de los cascos y otros recursos de hardware. Si bien existen experiencias con cascos de bajo costo como los Google Cardboard, las mismas suelen estar asociadas a entornos de educación no formal [2].

Los ámbitos de implementación de aplicaciones en RV se han ido ampliando a lo largo de los años. Se pueden encontrar experiencias de aplicación en medicina, telepresencia, empatía para tomar conciencia sobre situaciones críticas o situaciones que suceden en comunidades lejanas [3][4]. Se utiliza RV en el desarrollo de simuladores de vuelo, manejo de automóviles y otros vehículos, cuidado de personas con problemas médicos, además de los ámbitos más conocidos como educación y videojuegos. El uso de las características de la RV permite posicionarte en situaciones diferentes pero es importante tener en cuenta varios aspectos que hacen a la experiencia RV. Si bien el hardware y software es importante, no deja de ser menos relevante el estudio de la percepción humana para generar entornos realmente inmersivos [5].

En este contexto, se comenzó el estudio de técnicas interactivas de RV aplicadas a diferentes áreas de arquitectura que incluyen la revisión de diseños, la simulación de operaciones dinámicas, la coordinación del diseño detallado, la programación de la construcción y también para el marketing.

Es de interés para esta línea de investigación, analizar y aplicar técnicas de RV en el diseño arquitectónico y la construcción de tales diseños utilizando herramientas libres y en el ámbito local trabajando de manera interdisciplinaria con estudiantes y profesionales de la Facultad de Arquitectura y Urbanismo de nuestra ciudad.

Creemos que la RV es una herramienta altamente efectiva para presentar un diseño arquitectónico, pues permite transmitir adecuadamente las dimensiones espaciales, la información contextual y el realismo del diseño arquitectónico a los clientes.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

Si bien en el último años se avanzó en el área de RA y el desarrollo de aplicaciones con esta tecnología, también se comenzó a experimentar con Realidad Virtual y su aplicación en aplicaciones para profesionales del área de arquitectura, a fin de crear entornos que faciliten el uso de modelos arquitectónicos para su visualización y manipulación antes de la construcción. Esta aplicación se trabaja con el aporte de profesionales del área de Arquitectura y Urbanismo y, en un principio, está pensada para ser utilizada en la etapa de formación de los futuros arquitectos.

A continuación se detallan los ejes centrales de investigación:

- Relevamiento de herramientas para diseño 3D y renderización usadas por arquitectos (Sketchup¹, Lumion², V-Ray³).
- Análisis de motores de juegos existentes para RV (Unreal Engine 4, Unity, Godot, etc) a fin de determinar el apropiado para diferentes tipos de proyectos.
- Evaluación de distintos dispositivos dedicados para RV disponibles, analizando ventajas y desventajas de cada uno de ellos para determinar su adopción en el proyecto presentado.

¹ Sketchup: <https://www.sketchup.com/es/plans-and-pricing/sketchup-studio> Último acceso marzo 2022.

² Lumion: <https://www.lumion.es/> Último acceso marzo 2022.

³ V-Ray: <https://www.chaos.com/es/vray/sketchup> Último acceso marzo 2022.

3. RESULTADOS Y OBJETIVOS

El objetivo general de esa línea de trabajo es continuar con el uso de técnicas de RA orientadas a la educación y videojuegos serios y experimentar con RV interactiva en entornos arquitectónicos.

Para cumplir con el objetivo general, se proponen los siguientes objetivos específicos:

- Analizar herramientas para diseño 3D y renderización usadas por arquitectos y profesionales del área de la construcción.
- Continuar el desarrollo de aplicaciones interactivas que puedan utilizarse en el ámbito de la escuela y entornos de educación no formal, poniendo en práctica lo aprendido en el desarrollo con RA.
- Elaborar pautas de evaluación que permitan comprobar la adecuación y usabilidad de las aplicaciones realizadas con RA y RV.
- Determinar la efectividad de la tecnología de RV en la presentación de diseños arquitectónicos durante las etapas de revisión del mismo con el cliente de un proyecto de construcción. Para esta evaluación se tomará un proyecto real y local del sector de la construcción.
- Continuar promoviendo esta temática en otros escenarios y dentro del marco del desarrollo de las tesinas de grado.

Como resultado de esta línea de trabajo se ha publicado la experiencia de desarrollo y uso de la aplicación ERA, desarrollada en el marco de una tesina de grado y se ha puesto a disposición en la plataforma de distribución digital de aplicaciones móviles para los dispositivos con sistema

operativo Android operada por Google.

La tesina denominada **“ERA: entretenidos con realidad aumentada”** [6], fue el marco para el desarrollo de la aplicación móvil “Entretenidos con Kota” que acompaña al libro “Kota Corta” de Katia Maria Soto Kiewit, con arte de Tatiana Zanelli, como un paratexto del mismo⁴.

La Figura 1 muestra capturas de algunas de las escenas aumentadas del libro.

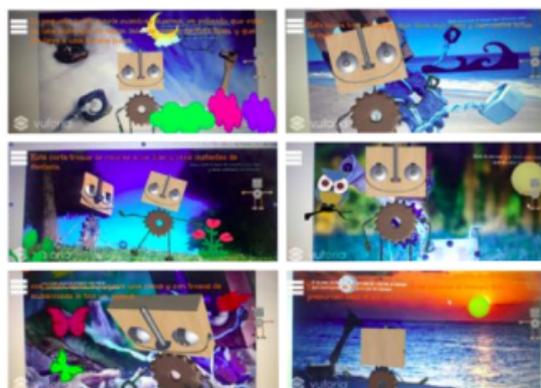


Figura 1. Escenas de aumentaciones del libro.

La Figura 2 ilustra la interacción con los aumentos provistos por la app.

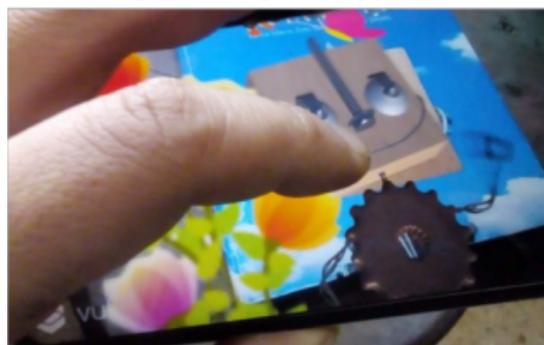


Figura 2. Experiencias de Entretenidos con Kota.

En el marco de este trabajo se definió una metodología para la evaluación de tecnologías relacionadas al desarrollo de aplicaciones con RA. En particular se definieron criterios que permiten evaluar

⁴ En este desarrollo se trabajó en conjunto con la cátedra Lenguaje Visual 3 de la Facultad de Artes de la UNLP, donde se elaboró el libro Kota Corta.

tanto los kits de desarrollo de RA, como los motores de juegos que permiten el desarrollo de la aplicación. Los SDK analizados fueron ARCore, ARToolkit, Wikitude, Vuforia, MAXST y Kudan. Los motores de juegos analizados fueron Unity, Unreal Engine 4, Godot y Panda 3D.

Entre los criterios definidos se puede mencionar:

- los mecanismos de reconocimiento y seguimiento de marcas para RA;
- la forma de realizar el aumento;
- tipo de licencia y soporte;
- plataformas o entornos de hardware y software en los que se encuentran disponibles;
- documentación disponible;
- la forma de integración de los SDK de RA con los motores de juego.

Las experiencias realizadas se llevaron a cabo en el Hospital Garrahan y en el Hospital de Niños de la ciudad de La Plata. Los resultados fueron publicados en la Conferencia Latinoamericana de Informática (CLEI) [7].

En cuanto a la línea de investigación de RV, se ha realizado un relevamiento de herramientas vinculadas a la creación de aplicaciones de RV, esto es, diseño 3D, renderizado y codificación de la app.

Asimismo se ha avanzado con el desarrollo del prototipo CEIT (“Creador de Entornos Inmersivos Transitables”), desarrollado para sistemas variables de Oculus RIFT que no requieren de dispositivos móviles. CEIT permite generar un entorno de RV a partir de un modelo 3D ingresado por el usuario.

La Figura 3 presenta una captura del tutorial interno provisto por CEIT donde se muestra un entorno que permite agregar y quitar objetos o experimentar

con texturas, a fin de que el usuario se familiarice con los controles del sistema.

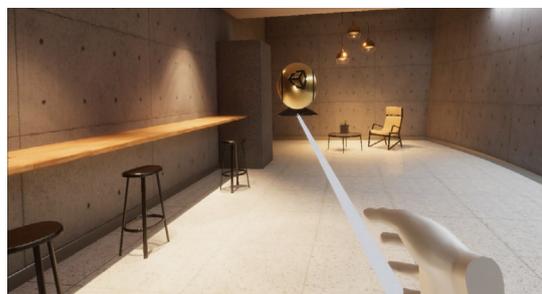


Figura 3. Captura de un entorno manipulable

Además CEIT proporciona funcionalidad para simular variaciones del ambiente relacionadas con cambios de horario, efectos climáticos y distintos tipos de iluminación. La Figura 4 muestra el impacto de estas variaciones en el modelo. En este caso, se puede seleccionar un horario y manipular la rotación del sol.

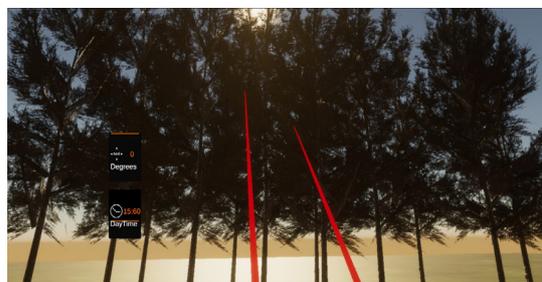


Figura 4. Captura de variación horaria, clima e iluminación

Esta aplicación además de generar espacios a partir de maquetas 3D, posibilita al usuario final recorrer e interactuar con ese espacio a través de la inmersión provista por la realidad virtual. Los trabajos presentados en este artículo se llevan a cabo de manera interdisciplinaria entre docentes y estudiantes de la Facultad de Informática, de Arquitectura y Urbanismo, y de Artes de la UNLP.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo de la línea de I+D+i presentada en este artículo está formado por docentes investigadores categorizados del LINTI y estudiantes de la Facultad de Informática. Se destaca la formación en equipos interdisciplinarios en los cuales participan docentes y profesionales de la Facultad de Artes y de la Facultad de Arquitectura y Urbanismo de la UNLP.

A través de la generación permanente de conocimiento por medio de líneas de investigación y desarrollo de aplicaciones vinculadas al sector productivo y su aplicación en el ámbito social, el LINTI promueve el uso innovador de las tecnologías informáticas en la región.

5. BIBLIOGRAFÍA

[1] Banchoff C, Fava, L., Schiavoni, A., Martin S. (2020). Realidad aumentada y realidad virtual aplicadas a proyectos con fines sociales. XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021). Chilecito. UNDeC. ISBN 978-987-24611-3-3.

[2] Chirinos Delfino, Y. (2020). La realidad virtual como mediadora de aprendizajes: desarrollo de una aplicación móvil de realidad virtual orientada a niños. Revista TE&ET no. 27. ISSN: 1850-9959. Páginas: 98-99.

[3] Craig, A, Sherman, B., William R., Jeffrey D. (2009). *Introduction to Virtual Reality, Developing Virtual Reality Applications*, Morgan Kaufmann, Boston, Pages 1-32, ISBN 978-0-12-374943-7, DOI: 0.1016/B978-0-12-374943-7.0000 1-X.

[4] Das, P.; Zhu, M.; McLaughlin, L.; Bilgrami, Z.; Milanaik, R.L. (2017). *Augmented reality video games: new possibilities and implications for children and adolescents*. Multimodal Technologies and Interaction, 2017; 1:8.

[5] M. Lavalley, M., *Virtual Reality*, 1st ed. Cambridge: Cambridge University Press, 2019. Disponible en: <http://vr.cs.uiuc.edu/vrbook.pdf>. Último acceso marzo de 2021.

[6] Jara, J. ERA: Entretenidos con Realidad Aumentada. Tesina de grado de la Licenciatura en Sistemas, Noviembre 2020. Disponible en <http://sedici.unlp.edu.ar/handle/10915/118525>

[7] J. D. Jara, C. Banchoff and L. Fava. Playful application with augmented reality in prolonged medical treatments. XLVII Latin American Computing Conference (CLEI), 2021, pp. 1-7, doi: 10.1109/CLEI53233.2021.9639948. Año 2021.